

```

!!SEQUENCE_LIST 1.0
! FINDPATTERNS on PIR: * allowing 0 mismatches
!
! <(X){1,200}(L,I,V,M,A,P)(X)(P,T,S)(L,I,V,M,A,P)(XX)(L,I,V,M,A,F,Y,W)(C,T,S)(R,K,H
!
PIR1:CCOF2T  ck: 2620  len: 92  finds: 1  | cytochrome c2 - Rhodocyclus re
PIR1:BDHCB  ck: 7216  len: 277  finds: 1  | carbonyl reductase (NADPH) (EC
PIR1:BDALAE ck: 2829  len: 246  finds: 1  | acetoacetyl-CoA reductase (EC
PIR1:S05397  ck: 3104  len: 272  finds: 1  | granaticin polyketide ketoredu
PIR1:S25079  ck: 4473  len: 261  finds: 1  | monensin polyketide ketoreduct
PIR1:BA2147  ck: 3492  len: 244  finds: 1  | 3-oxoacyl-[acyl-carrier-prote
PIR1:D69930  ck: 2171  len: 238  finds: 1  | probable 3-oxoacyl-[acyl-carri
PIR1:RDECEP  ck: 6342  len: 334  finds: 1  | N-acetyl-gamma-glutamyl-phosph
PIR1:DTECC  ck: 4302  len: 311  finds: 1  | aspartate carboxyltransferase
PIR1:OWEBAC  ck: 4073  len: 311  finds: 1  | aspartate carboxyltransferase
PIR1:OWSEAC  ck: 6543  len: 306  finds: 1  | aspartate carboxyltransferase
PIR1:OWZP  ck: 9240  len: 327  finds: 1  | ornithine carboxyltransferase
PIR1:A69962  ck: 158  len: 299  finds: 1  | phosphate butyryltransferase
PIR1:XNEBHC  ck: 1849  len: 359  finds: 1  | histidinol-phosphate transamin
PIR1:S29090  ck: 5401  len: 367  finds: 1  | dual specificity phosphoprotei
PIR1:S24411  ck: 5871  len: 367  finds: 1  | dual specificity phosphoprotei
PIR1:PRMSCL  ck: 92  len: 247  finds: 1  | cytotoxic T-lymphocyte protein
PIR1:OKSE  ck: 107  len: 376  finds: 1  | beta-lactamase (EC 3.5.2.6) pr
PIR1:DS3312  ck: 6718  len: 133  finds: 1  | cytidine deaminase (EC 3.5.4.5
PIR1:JE0022  ck: 9048  len: 136  finds: 1  | cytidine/deoxycytidine deamin
PIR1:F69500  ck: 4460  len: 219  finds: 1  | riboflavin-specific deaminase
PIR1:I36612  ck: 2178  len: 142  finds: 1  | H+-transporting ATP synthase
PIR1:RKIOS2  ck: 4535  len: 180  finds: 1  | ribulose-bisphosphate carboxyl
PIR1:RKIOS1  ck: 5240  len: 181  finds: 1  | ribulose-bisphosphate carboxyl
PIR1:RKPOS1  ck: 4699  len: 181  finds: 1  | ribulose-bisphosphate carboxyl
PIR1:RKPOS1  ck: 4863  len: 181  finds: 1  | ribulose-bisphosphate carboxyl
PIR1:RKPOS1  ck: 2967  len: 180  finds: 1  | ribulose-bisphosphate carboxyl
PIR1:F69398  ck: 4436  len: 277  finds: 1  | dihydroxynaphthoic acid syntha
PIR1:F71428  ck: 9250  len: 244  finds: 1  | hypothetical protein - Arabid
PIR1:A48681  ck: 300  len: 376  finds: 1  | placental thrombin inhibitor
PIR1:WIWTA  ck: 4689  len: 123  finds: 1  | alpha-amylase inhibitor cIII
PIR1:WIWTA5  ck: 4073  len: 124  finds: 1  | alpha-amylase inhibitor 0.53
PIR1:WIWTA1  ck: 3320  len: 124  finds: 1  | alpha-amylase inhibitor 0.19
PIR1:TVHUP1  ck: 4677  len: 380  finds: 1  | transforming protein fos - hum
PIR1:TVCHNS  ck: 7007  len: 367  finds: 1  | transforming protein fos -
PIR1:TVFEF4  ck: 9324  len: 322  finds: 1  | transforming protein fos -
PIR1:S00755  ck: 1536  len: 350  finds: 1  | pleckstrin - human
PIR1:HLHUG  ck: 908  len: 216  finds: 1  | class II histocompatibility
PIR1:RLJLV8  ck: 6002  len: 132  finds: 1  | ribosomal protein S8 - live
PIR1:S26296  ck: 6968  len: 151  finds: 1  | ribosomal protein S13-e, cy
PIR1:A48103  ck: 718  len: 256  finds: 1  | ribosomal protein L21 precu
PIR1:QJ298  ck: 409  len: 340  finds: 1  | annexin II type 2 - African
PIR1:OOFF  ck: 4358  len: 373  finds: 1  | rhodopsin - fruit fly (Dros
PIR1:I55604  ck: 4750  len: 411  finds: 1  | platelet glycoprotein Ib be
PIR1:QREBOB  ck: 4920  len: 306  finds: 1  | oligopeptide transport syst
PIR1:S67566  ck: 8415  len: 417  finds: 1  | probable membrane protein y
PIR1:S30896  ck: 7059  len: 297  finds: 1  | virulence-associated protei
PIR1:S06670  ck: 7032  len: 297  finds: 1  | virulence-associated protei
PIR1:IMECNA  ck: 6528  len: 131  finds: 1  | colicin N immunity protein
PIR1:ZMECD  ck: 8403  len: 139  finds: 1  | proteinase unmd (EC 3.4.21.
PIR1:HE4563  ck: 9038  len: 343  finds: 1  | flagellar motor switch prot
PIR1:BVECU  ck: 4271  len: 130  finds: 1  | unci protein - Escherichia
PIR1:BBAG55  ck: 173  len: 257  finds: 1  | virB8 protein precursor - A
PIR1:BBAG6  ck: 2646  len: 230  finds: 1  | virB8 protein - Agrobacteri
PIR1:BBAG58  ck: 6263  len: 237  finds: 1  | virB8 protein - Agrobacteri
PIR1:W2WL31  ck: 9656  len: 372  finds: 1  | E2 protein - human papillom
PIR1:W2WL51  ck: 1816  len: 358  finds: 1  | E2 protein - human papillom
PIR1:W2WL18  ck: 5567  len: 365  finds: 1  | E2 protein - human papillom
PIR1:SI5617  ck: 734  len: 391  finds: 1  | E2 protein - human papillom
PIR1:QOBENS  ck: 7071  len: 366  finds: 1  | integral membrane protein -
PIR1:WMNZ22  ck: 9828  len: 194  finds: 1  | envelope-associated 22k pro
PIR1:WMNZBA  ck: 6337  len: 186  finds: 1  | matrix glycoprotein M2 - bo
PIR1:WMVYP2  ck: 3415  len: 109  finds: 1  | 12k protein - potato virus
PIR1:VCBWC  ck: 3595  len: 237  finds: 1  | coat protein - tobacco stre
PIR1:VCVETC  ck: 1818  len: 351  finds: 1  | coat protein - turnip crink
PIR1:VHBPF6  ck: 6252  len: 149  finds: 1  | major nucleocapsid protein
PIR1:ZGBG4  ck: 4484  len: 177  finds: 1  | gene G protein - phage G4
PIR1:PDGYC  ck: 6596  len: 296  finds: 1  | REP2 protein - yeast (Sacch
PIR1:BE4166  ck: 1216  len: 214  finds: 1  | hypothetical protein H1084
PIR1:E70044  ck: 4650  len: 310  finds: 1  | conserved hypothetical prot

```

PIR1:G69000	ck: 7972	len: 409	finds: 1	molybdenum cofactor biosynthes
PIR1:E70191	ck: 8273	len: 337	finds: 1	conserved hypothetical protein
PIR1:A71183	ck: 8100	len: 306	finds: 1	probable UDP-glucose 4-epimerase
PIR1:S70955	ck: 8920	len: 288	finds: 1	ctnf protein - Vldrl cholesterol
PIR1:A70118	ck: 6242	len: 299	finds: 1	glycine betaine, L-proline ABC
PIR2:S76375	ck: 6628	len: 126	finds: 1	plastocyanin precursor - Synec
PIR2:F70646	ck: 8411	len: 323	finds: 1	probable quinone oxidoreductase
PIR2:S01901	ck: 8887	len: 254	finds: 1	alcohol dehydrogenase (EC 1.1.
PIR2:S65732	ck: 4453	len: 132	finds: 1	carbonyl reductase (NADPH) (EC
PIR2:SC5284	ck: 7185	len: 277	finds: 1	carbonyl reductase (NADPH) (EC
PIR2:SS4815	ck: 5190	len: 261	finds: 1	urdamycin polyketide ketoreduc
PIR2:D55587	ck: 3920	len: 261	finds: 1	griseusin polyketide ketoreduc
PIR2:S69225	ck: 7076	len: 262	finds: 1	nogalamycin polyketide ketored
PIR2:S35196	ck: 9329	len: 261	finds: 1	hypothetical protein 3 - Sacch
PIR2:D64051	ck: 4159	len: 242	finds: 1	3-oxoacyl-[acyl-carrier-prote
PIR2:D69648	ck: 3484	len: 254	finds: 1	2-keto-3-deoxygluconate oxidor
PIR2:C48674	ck: 5293	len: 268	finds: 1	tropinone reductase homolog -
PIR2:T12020	ck: 9922	len: 247	finds: 1	3-oxoacyl-acyl carrier protein
PIR2:T12051	ck: 5984	len: 244	finds: 1	3-oxoacyl-[acyl-carrier-prote
PIR2:S39654	ck: 5853	len: 255	finds: 1	hypothetical protein X - Pseud
PIR2:H69885	ck: 6790	len: 242	finds: 1	3-oxoacyl-acyl-carrier prote
PIR2:E69686	ck: 7200	len: 248	finds: 1	glucose 1-dehydrogenase homo
PIR2:BS5850	ck: 1047	len: 267	finds: 1	hypothetical protein (hsdh 5'
PIR2:A56275	ck: 101	len: 387	finds: 1	1,3-propanediol dehydrogenase
PIR2:S47829	ck: 4443	len: 339	finds: 1	glycerol-3-phosphate dehydroge
PIR2:A44132	ck: 6742	len: 338	finds: 1	D-aspartate oxidase (EC 1.4.3.
PIR2:SC5438	ck: 7167	len: 341	finds: 1	D-aspartate oxidase (EC 1.4.3.
PIR2:SC5439	ck: 934	len: 282	finds: 1	D-aspartate oxidase (EC 1.4.3.
PIR2:S25422	ck: 6471	len: 72	finds: 1	NADH dehydrogenase (ubiquinone
PIR2:S08622	ck: 675	len: 307	finds: 1	NADH dehydrogenase (ubiquinone
PIR2:T11338	ck: 2220	len: 347	finds: 1	NADH dehydrogenase (ubiquinone
PIR2:T11159	ck: 979	len: 114	finds: 1	NADH dehydrogenase 3 - hardbac
PIR2:S42244	ck: 1655	len: 214	finds: 1	NADH dehydrogenase (ubiquinone
PIR2:F64999	ck: 2909	len: 184	finds: 1	NADH dehydrogenase (ubiquinone
PIR2:S52970	ck: 6545	len: 167	finds: 1	NADH dehydrogenase (ubiquinone
PIR2:S78138	ck: 8115	len: 267	finds: 1	cytochrome-c oxidase (EC 1.9.3
PIR2:S07557	ck: 7659	len: 268	finds: 1	cytochrome-c oxidase (EC 1.9.3
PIR2:T09997	ck: 1856	len: 385	finds: 1	lipoxigenase (EC 1.13.11.12
PIR2:S43771	ck: 2848	len: 347	finds: 1	phosphatidylethanolamine desatur
PIR2:S74692	ck: 6210	len: 231	finds: 1	tRNA(mIG37)methyltransferase
PIR2:A71119	ck: 9980	len: 308	finds: 1	probable aspartate carbamoy
PIR2:A70959	ck: 1547	len: 319	finds: 1	probable aspartate carbamoy
PIR2:C70815	ck: 556	len: 403	finds: 1	probable beta-ketoadipyl Co
PIR2:T12981	ck: 8558	len: 452	finds: 1	glucuronosyl transferase-11
PIR2:T140986	ck: 7621	len: 323	finds: 1	streptomycin/spectinomycin
PIR2:I65760	ck: 800	len: 285	finds: 1	alcohol sulfotransferase (E
PIR2:S52265	ck: 6443	len: 367	finds: 1	dual specificity phosphatas
PIR2:T12078	ck: 3061	len: 231	finds: 1	ribonuclease (EC 3.1.-.-) -
PIR2:B71518	ck: 3917	len: 167	finds: 1	probable lipoprotein signal
PIR2:S26429	ck: 9885	len: 178	finds: 1	probable dUTP pyrophosphata
PIR2:S42592	ck: 4170	len: 131	finds: 1	urease (EC 3.5.1.5), tissue
PIR2:I51663	ck: 4321	len: 360	finds: 1	arginase 1 - African clawed
PIR2:I51664	ck: 4641	len: 360	finds: 1	arginase 2 - African clawed
PIR2:I51665	ck: 4724	len: 360	finds: 1	arginase 3 - African clawed
PIR2:J06069	ck: 8425	len: 140	finds: 1	blastididin-S deaminase (EC
PIR2:S37304	ck: 2135	len: 335	finds: 1	spal protein - Salmonella t
PIR2:S01397	ck: 4392	len: 127	finds: 1	H+-transporting ATP synthas
PIR2:S22348	ck: 1686	len: 168	finds: 1	H+-transporting ATP synthas
PIR2:D72121	ck: 2102	len: 209	finds: 1	ATP synthase chain D - Chia
PIR2:S31498	ck: 4433	len: 181	finds: 1	ribulose-bisphosphate carbo
PIR2:A71295	ck: 255	len: 357	finds: 1	probable alanine racemase (
PIR2:G70553	ck: 8877	len: 268	finds: 1	probable echA10 protein - M
PIR2:H69457	ck: 4102	len: 322	finds: 1	ornithine cyclodeaminase (a
PIR2:S75089	ck: 8968	len: 338	finds: 1	UDP-glucose 4-epimerase - S
PIR2:H69105	ck: 3451	len: 356	finds: 1	dUDP-glucose 4,6-dehydratas
PIR2:D65290	ck: 128	len: 332	finds: 1	dUDP-glucose 4,6-dehydratas
PIR2:S47045	ck: 5086	len: 346	finds: 1	dUDP-glucose 4,6-dehydratas
PIR2:E70566	ck: 5640	len: 331	finds: 1	probable DTDP-GLUCOSE 4 - M
PIR2:G70415	ck: 2675	len: 321	finds: 1	nucleotide sugar epimerase
PIR2:E69750	ck: 3388	len: 249	finds: 1	glucosamine-6-phosphate iso
PIR2:D69094	ck: 5550	len: 248	finds: 1	phosphoribosylaminimidazole
PIR2:A19940	ck: 1681	len: 40	finds: 1	antithrombin III - baboon (
PIR2:JX0206	ck: 7697	len: 207	finds: 1	chymotrypsin inhibitor (Kun

PIR2:JS0650	CK: 8167	len: 207	finds: 1	! chymotrypsin inhibitor (Kunitz
PIR2:JX0246	CK: 1378	len: 181	finds: 1	! serine proteinase inhibitor A
PIR2:JC5447	CK: 4632	len: 210	finds: 1	! serine proteinase inhibitor A
PIR2:TO6517	CK: 7537	len: 151	finds: 1	! alpha-amylase inhibitor Ima1 F
PIR2:SI6920	CK: 2886	len: 121	finds: 1	! alpha-amylase inhibitor - whea
PIR2:A60195	CK: 2543	len: 278	finds: 1	! transforming protein dbl - mou
PIR2:AA0316	CK: 7358	len: 415	finds: 1	! cyclin B - fission yeast (Schl
PIR2:AS7234	CK: 2377	len: 348	finds: 1	! lin-44 protein precursor - Cae
PIR2:JC4152	CK: 4630	len: 354	finds: 1	! Wnt-11 protein precursor - chi
PIR2:S34378	CK: 1689	len: 354	finds: 1	! wnt-11 protein - mouse
PIR2:JH0688	CK: 1620	len: 398	finds: 1	! bone morphogenetic protein 211
PIR2:JH0687	CK: 9563	len: 398	finds: 1	! bone morphogenetic protein 21
PIR2:S21299	CK: 8703	len: 81	finds: 1	! protein Daf8 - eastern green m
PIR2:AS7045	CK: 6756	len: 189	finds: 1	! thermostable direct hemolysin
PIR2:AA8942	CK: 5895	len: 189	finds: 1	! thermostable direct hemolysin
PIR2:JQ0472	CK: 2831	len: 135	finds: 1	! T-cell receptor beta chain (BT
PIR2:JQ0473	CK: 3911	len: 136	finds: 1	! T-cell receptor beta chain (BT
PIR2:BS5250	CK: 6952	len: 353	finds: 1	! class I histocompatibility ant
PIR2:IS0609	CK: 8972	len: 207	finds: 1	! T-cell surface glycoprotein CD
PIR2:IS3916	CK: 7310	len: 257	finds: 1	! ribosomal protein S3 - Achnola
PIR2:SS77489	CK: 8926	len: 133	finds: 1	! ribosomal protein S8 - Synech
PIR2:T01947	CK: 4063	len: 252	finds: 1	! hypothetical protein Fllo4.1 -
PIR2:AS9234	CK: 9791	len: 371	finds: 1	! opsin - bluebottle fly (Calli
PIR2:SS40691	CK: 7181	len: 374	finds: 1	! opsin rh1 - fruit fly (Drosoph
PIR2:JC4304	CK: 2863	len: 355	finds: 1	! orphan G-protein-coupled rece
PIR2:IS58186	CK: 4478	len: 354	finds: 1	! hypothetical G-protein coupled
PIR2:AS29667	CK: 8980	len: 60	finds: 1	! pulmonary surfactant protein B
PIR2:SS22494	CK: 429	len: 277	finds: 1	! rRNA N-glycosidase (EC 3.2.2.2
PIR2:SS23519	CK: 9157	len: 278	finds: 1	! beta-luflin - smooth loofah
PIR2:JN0108	CK: 7681	len: 250	finds: 1	! luflin-b - smooth loofah
PIR2:SS65052	CK: 5285	len: 214	finds: 1	! pistill-specific protein sts14
PIR2:JC4822	CK: 8148	len: 77	finds: 1	! acyl carrier protein - Bacillus
PIR2:BS65003	CK: 7651	len: 228	finds: 1	! histidine transport system per
PIR2:BS6263	CK: 5147	len: 306	finds: 1	! oligopeptide transport system
PIR2:C71163	CK: 4618	len: 304	finds: 1	! probable oligopeptide transport
PIR2:SS64728	CK: 6546	len: 482	finds: 1	! protein secretion protein xcp1
PIR2:SS2163	CK: 8969	len: 331	finds: 1	! sucrose specific repressor - B
PIR2:D70044	CK: 4428	len: 243	finds: 1	! transcription regulator Gnt
PIR2:SS48826	CK: 9647	len: 215	finds: 1	! NADH dehydrogenase (ubiquin
PIR2:SS75086	CK: 4153	len: 342	finds: 1	! iron-stress chlorophyll-b1n
PIR2:AS64334	CK: 3224	len: 182	finds: 1	! probable fibinrial protein-1
PIR2:D71484	CK: 1871	len: 179	finds: 1	! probable ribosome releasing
PIR2:SS74877	CK: 7310	len: 328	finds: 1	! phosphate transport system
PIR2:AA6259	CK: 9097	len: 438	finds: 1	! recombination protein reca
PIR2:JQ0661	CK: 6434	len: 424	finds: 1	! impb protein - Salmonella t
PIR2:ET1946	CK: 9168	len: 343	finds: 1	! flagellar motor switch prot
PIR2:AS37763	CK: 2312	len: 147	finds: 1	! virD1 protein - Agrobacteri
PIR2:AS25063	CK: 4145	len: 147	finds: 1	! hypothetical virD1 protein
PIR2:SS06883	CK: 2761	len: 147	finds: 1	! virD1 protein - Agrobacteri
PIR2:SS6564	CK: 5828	len: 368	finds: 1	! E2 protein - human papillom
PIR2:SS6576	CK: 4491	len: 368	finds: 1	! E2 protein - human papillom
PIR2:SS6552	CK: 4819	len: 383	finds: 1	! E2 protein - human papillom
PIR2:SS6535	CK: 273	len: 376	finds: 1	! E2 protein - human papillom
PIR2:AA4215	CK: 5996	len: 303	finds: 1	! EUS1 protein - equine herpe
PIR2:FA3674	CK: 501	len: 372	finds: 1	! US7 protein - human herpesv
PIR2:SS8346	CK: 2806	len: 256	finds: 1	! coat protein V1 - tomato ye
PIR2:Q1248	CK: 566	len: 106	finds: 1	! hypothetical 12K protein -
PIR2:SS41285	CK: 3148	len: 355	finds: 1	! coat protein - sweet potato
PIR2:ET1669	CK: 3620	len: 433	finds: 1	! preprotein translocase secY
PIR2:CG9632	CK: 4284	len: 258	finds: 1	! transcription repressor glc
PIR2:F69260	CK: 6950	len: 244	finds: 1	! nitrate ABC transporter, pe
PIR2:A70001	CK: 6987	len: 253	finds: 1	! ABC transporter (ATP-bindin
PIR2:AS70039	CK: 1714	len: 301	finds: 1	! ABC transporter (ATP-bindin
PIR2:SS69260	CK: 9269	len: 243	finds: 1	! nitrate ABC transporter, AT
PIR2:SS6325	CK: 1039	len: 252	finds: 1	! hisP-like nucleotide bindin
PIR2:SS73545	CK: 2944	len: 450	finds: 1	! ATP-dependent RNA helicase
PIR2:SS69260	CK: 2958	len: 482	finds: 1	! probable RNA helicase CAS/6
PIR2:AS38900	CK: 3980	len: 272	finds: 1	! T-cell-specific transcript1
PIR2:JH0402	CK: 9457	len: 279	finds: 1	! T-cell-specific transcript1
PIR2:SS0068	CK: 241	len: 393	finds: 1	! nonhistone chromosomal prot
PIR2:AS38095	CK: 5007	len: 269	finds: 1	! T-cell-specific transcript1
PIR2:BS38095	CK: 3201	len: 269	finds: 1	! T-cell-specific transcript1
PIR2:DS38095	CK: 7876	len: 250	finds: 1	! T-cell-specific transcript1

PIR2:CC8095	ck: 409	len: 268	finds: 1	T-cell-specific transcription	PIR2:D64151	ck: 1410	len: 430	finds: 1	hypothetical protein H10404
PIR2:JC6179	ck: 83	len: 393	finds: 1	dorsal switch protein 1 - fruit	PIR2:S75967	ck: 3587	len: 271	finds: 1	hypothetical protein - Sme
PIR2:SS1796	ck: 5713	len: 366	finds: 1	T-cell-specific transcription	PIR2:D69014	ck: 6260	len: 222	finds: 1	conserved hypothetical prot
PIR2:JN0767	ck: 3795	len: 270	finds: 1	homeobox protein HEX - human	PIR2:CG5012	ck: 4986	len: 285	finds: 1	hypothetical protein b2382
PIR2:I46089	ck: 4988	len: 371	finds: 1	thyroid transcription factor-1	PIR2:A70606	ck: 144	len: 358	finds: 1	probable electron transfer
PIR2:SS0230	ck: 8616	len: 271	finds: 1	homeotic protein Hex - mouse	PIR2:CG5112	ck: 597	len: 220	finds: 1	sigma cross-reacting protei
PIR2:SS3724	ck: 7622	len: 372	finds: 1	thyroid-specific enhancer-bin	PIR2:JC5603	ck: 1627	len: 270	finds: 1	Esi protein - zebra fish
PIR2:AS6451	ck: 5439	len: 371	finds: 1	thyroid-specific enhancer-bind	PIR2:T09924	ck: 1428	len: 293	finds: 1	cytidine deaminase (EC 3.5.
PIR2:SS12002	ck: 6608	len: 372	finds: 1	thyroid nuclear factor 1 - rat	PIR2:SS5551	ck: 2728	len: 257	finds: 1	cytokine-inducible protein
PIR2:IS9234	ck: 9010	len: 420	finds: 1	octamer binding transcription	PIR2:JC5626	ck: 1738	len: 198	finds: 1	STAT induced STAT inhibitor
PIR2:G02321	ck: 1141	len: 401	finds: 1	thyroid transcription factor 1	PIR2:JC5760	ck: 1381	len: 198	finds: 1	cytokine-inducible SR2 prot
PIR2:S78452	ck: 3745	len: 331	finds: 1	POU-domain protein rdc-1 - hum	PIR2:A41900	ck: 4304	len: 299	finds: 1	cyn operon regulatory prote
PIR2:S76329	ck: 4655	len: 416	finds: 1	probable phosphoesterase (EC 3	PIR2:F70721	ck: 7074	len: 164	finds: 1	hypothetical protein RV1829
PIR2:F71508	ck: 6469	len: 244	finds: 1	probable phosphoesterase (EC 3	PIR2:SS6618	ck: 3873	len: 173	finds: 1	yjix protein - Escherichia
PIR2:SS3782	ck: 5295	len: 467	finds: 1	serine/threonine protein kinase	PIR2:S45255	ck: 4167	len: 54	finds: 1	yjix protein homolog - Ente
PIR2:CG9230	ck: 3724	len: 257	finds: 1	conserved hypothetical protein	PIR2:A70738	ck: 8373	len: 158	finds: 1	probable rimi protein - Myc
PIR2:AA0369	ck: 4295	len: 302	finds: 1	trans-acting transcription reg	PIR2:F72316	ck: 5138	len: 309	finds: 1	hypothetical protein TM0917
PIR2:AA3960	ck: 1522	len: 306	finds: 1	cytochrome caa3 oxidase ctaa -	PIR2:B70451	ck: 4885	len: 311	finds: 1	prephenate dehydrogenase -
PIR2:AA5335	ck: 368	len: 297	finds: 1	cytochrome-c oxidase assembly	PIR2:G39845	ck: 2658	len: 256	finds: 1	dihydroorotate dehydrogenas
PIR2:G02514	ck: 2327	len: 326	finds: 1	P2 purinoceptor - human	PIR2:JC01382	ck: 6785	len: 303	finds: 1	hypothetical 3A4 protein -
PIR2:JC4800	ck: 7545	len: 328	finds: 1	P216 receptor - human	PIR2:SS6961	ck: 8234	len: 196	finds: 1	probable membrane protein Y
PIR2:FF4471	ck: 3696	len: 415	finds: 1	hypothetical protein homolog M	PIR2:JC01355	ck: 9413	len: 134	finds: 1	v0 protein - Muscathus str
PIR2:AA7201	ck: 1219	len: 120	finds: 1	bombesinlike peptide - African	PIR2:JC0150	ck: 6657	len: 122	finds: 1	hypothetical 13K protein -
PIR2:EE5139	ck: 1678	len: 197	finds: 1	hypothetical protein B3434 - E	PIR2:B71217	ck: 8463	len: 108	finds: 1	hypothetical protein PH2002
PIR2:SA48737	ck: 6122	len: 423	finds: 1	kynurenine aminotransferase -	PIR2:T05820	ck: 4391	len: 217	finds: 1	hypothetical protein TSK18.
PIR2:JC02268	ck: 6645	len: 364	finds: 1	O-methyltransferase (EC 2.1.1.	PIR2:S77462	ck: 3256	len: 275	finds: 1	hypothetical protein s11107
PIR2:G70165	ck: 3950	len: 315	finds: 1	aldose reductase homolog - Lym	PIR2:SS3572	ck: 4181	len: 158	finds: 1	probable membrane protein Y
PIR2:G69076	ck: 4836	len: 251	finds: 1	molybdopterin biosynthesis pro	PIR2:GA5335	ck: 1335	len: 146	finds: 1	ORF1 protein - Autographa c
PIR2:SS6490	ck: 2843	len: 332	finds: 1	hypothetical transcription reg	PIR2:SS2975	ck: 4975	len: 383	finds: 1	gene BCR2 protein - human
PIR2:HE9823	ck: 1929	len: 302	finds: 1	conserved hypothetical protein	PIR2:SO2383	ck: 5478	len: 57	finds: 1	probable membrane antigen C
PIR2:AA4223	ck: 590	len: 308	finds: 1	probable pseudouridine synthas	PIR2:SO3243	ck: 3042	len: 115	finds: 1	hypothetical protein B-115
PIR2:S73869	ck: 6533	len: 309	finds: 1	hypothetical protein ycec - My	PIR2:SS21504	ck: 7782	len: 148	finds: 1	hypothetical protein (gene
PIR2:BB4026	ck: 2974	len: 164	finds: 1	hypothetical protein H11338 -	PIR2:SS5282	ck: 7758	len: 317	finds: 1	eaa protein - phage P22
PIR2:BB5007	ck: 9360	len: 161	finds: 1	hypothetical protein b2340 - E	PIR2:SI7834	ck: 1383	len: 36	finds: 1	acetyl-CoA carboxylase - ph
PIR2:CG4882	ck: 51	len: 299	finds: 1	probable transcription regulat	PIR2:AS6663	ck: 6051	len: 299	finds: 1	capsid protein VPI - foot-a
PIR2:ET2058	ck: 4148	len: 202	finds: 1	probable phosphatase/kinase -	PIR2:HT2245	ck: 6860	len: 236	finds: 1	2,3,4,5-tetrahydropyridine-
PIR2:BV1509	ck: 4232	len: 202	finds: 1	probable phosphatase/kinase -					



PIR2:E72316	ck: 6581	len: 410	finds: 1	i	hypothetical protein - Thermot
PIR2:BA0609	ck: 8809	len: 236	finds: 1	i	probable lipoprotein Mlpa - My
PIR2:JC6019	ck: 2747	len: 290	finds: 1	i	response-regulator protein Fir
PIR2:CT0198	ck: 1619	len: 440	finds: 1	i	conserved hypothetical protein
PIR2:E70167	ck: 2658	len: 192	finds: 1	i	hypothetical protein BB0542 -
PIR2:B70174	ck: 8103	len: 208	finds: 1	i	hypothetical protein BB0595 -
PIR2:D71926	ck: 2840	len: 142	finds: 1	i	cag island protein - Helicobac
PIR2:F71925	ck: 6963	len: 114	finds: 1	i	cag island protein - Helicobac
PIR2:H64586	ck: 7347	len: 114	finds: 1	i	cag pathogenicity island prote
PIR2:F64587	ck: 2551	len: 142	finds: 1	i	cag pathogenicity island prote
PIR2:A64649	ck: 6457	len: 131	finds: 1	i	hypothetical protein HP1033 -
PIR2:D64710	ck: 2117	len: 115	finds: 1	i	hypothetical protein HP1524 -
PIR2:E71938	ck: 3806	len: 91	finds: 1	i	hypothetical protein jhp0391 -
PIR2:A71809	ck: 3881	len: 115	finds: 1	i	hypothetical protein jhp1413 -
PIR2:C71912	ck: 428	len: 307	finds: 1	i	probable outer membrane protea
PIR2:PH0856	ck: 7474	len: 277	finds: 1	i	mauf protein - Paracoccus deni
PIR2:I19703	ck: 4597	len: 102	finds: 1	i	tram protein - Agrobacterium t
PIR2:JC5568	ck: 7036	len: 118	finds: 1	i	mrxy protein - Rhizobium meli
PIR2:S27344	ck: 1548	len: 370	finds: 1	i	hupK protein - Rhizobium legum
PIR2:S28677	ck: 6944	len: 243	finds: 1	i	hypothetical protein 4 - Rhizoc
PIR2:S26139	ck: 4831	len: 187	finds: 1	i	signalling protein ampD - Citr
PIR2:AA8901	ck: 4831	len: 187	finds: 1	i	signalling protein ampD - Ente
PIR2:S40867	ck: 3905	len: 248	finds: 1	i	ferredoxin--NADP+ reductase (E
PIR2:A64750	ck: 8141	len: 79	finds: 1	i	hypothetical protein b0249 - E
PIR2:C65059	ck: 1867	len: 425	finds: 1	i	hypothetical protein b2775 - E
PIR2:B65068	ck: 6883	len: 145	finds: 1	i	hypothetical protein b2849 - E
PIR2:F64963	ck: 7029	len: 359	finds: 1	i	nicotinate-nucleotide--dimethyl
PIR2:T00211	ck: 5937	len: 291	finds: 1	i	Type II secretion pathway rela
PIR2:H64726	ck: 98	len: 216	finds: 1	i	yabP protein - Escherichia coli
PIR2:A64862	ck: 1943	len: 78	finds: 1	i	ycgZ protein - Escherichia coli
PIR2:A64882	ck: 6021	len: 262	finds: 1	i	ycjI protein - Escherichia coli
PIR2:S20452	ck: 9681	len: 271	finds: 1	i	hypothetical protein X - Klebs
PIR2:S01838	ck: 7149	len: 220	finds: 1	i	nifH protein - Klebsiella pneu
PIR2:S22619	ck: 7734	len: 336	finds: 1	i	hypothetical protein - Salmon
PIR2:AS3302	ck: 5663	len: 289	finds: 1	i	hemf 5'-region hypothetical pr
PIR2:S23906	ck: 4054	len: 135	finds: 1	i	hypothetical protein 1 - Shiga
PIR2:TI4663	ck: 2204	len: 99	finds: 1	i	histone H5 like protein - Yers
PIR2:TI4952	ck: 2530	len: 144	finds: 1	i	hypothetical protein - Yers
PIR2:S70883	ck: 2492	len: 227	finds: 1	i	hypothetical protein 3 - Vi
PIR2:S54442	ck: 6509	len: 150	finds: 1	i	membrane associated protein
PIR2:G64102	ck: 8973	len: 288	finds: 1	i	phosphatidate cytidyltran
PIR2:H64157	ck: 4	len: 311	finds: 1	i	suif protein homolog HI0733
PIR2:H71692	ck: 1607	len: 405	finds: 1	i	hypothetical protein RP358
PIR2:E72060	ck: 2665	len: 252	finds: 1	i	cta70 hypothetical protein
PIR2:F72031	ck: 6017	len: 493	finds: 1	i	cta78 hypothetical protein
PIR2:S61492	ck: 5418	len: 270	finds: 1	i	hypothetical protein 2 - Ch
PIR2:JC5205	ck: 8587	len: 160	finds: 1	i	sulfur-rich protein - Chlam
PIR2:A71529	ck: 4189	len: 303	finds: 1	i	hypothetical protein CT324
PIR2:C71569	ck: 1406	len: 300	finds: 1	i	probable ribonuclease HII -
PIR2:B53203	ck: 1180	len: 43	finds: 1	i	hypothetical protein 2 - De
PIR2:S17812	ck: 2127	len: 192	finds: 1	i	hypothetical protein 192 -
PIR2:S22631	ck: 1875	len: 166	finds: 1	i	transcription negative regu
PIR2:B58883	ck: 8138	len: 213	finds: 1	i	hypothetical protein 7 (atp
PIR2:S36981	ck: 4069	len: 201	finds: 1	i	heme binding protein precur
PIR2:S74634	ck: 2111	len: 393	finds: 1	i	hypothetical protein - syne
PIR2:S76728	ck: 9759	len: 187	finds: 1	i	hypothetical protein - syne
PIR2:S76937	ck: 9556	len: 312	finds: 1	i	hypothetical protein - syne
PIR2:S74351	ck: 4892	len: 333	finds: 1	i	hypothetical protein sll007
PIR2:S74733	ck: 376	len: 156	finds: 1	i	hypothetical protein sll093
PIR2:S74723	ck: 8425	len: 125	finds: 1	i	hypothetical protein sll093
PIR2:S75160	ck: 4206	len: 215	finds: 1	i	hypothetical protein sll163
PIR2:S77220	ck: 3780	len: 285	finds: 1	i	hypothetical protein sll144
PIR2:S77428	ck: 4795	len: 79	finds: 1	i	hypothetical protein srr206
PIR2:B41838	ck: 3600	len: 384	finds: 1	i	Vans - Enterococcus faecium
PIR2:S69337	ck: 3316	len: 224	finds: 1	i	conserved hypothetical prot
PIR2:S39739	ck: 3644	len: 394	finds: 1	i	efflux protein homolog ywfF
PIR2:C42365	ck: 1150	len: 208	finds: 1	i	flagellar assembly protein
PIR2:C69820	ck: 5172	len: 83	finds: 1	i	hypothetical protein yhaY -
PIR2:S69876	ck: 1724	len: 36	finds: 1	i	hypothetical protein yrkG -
PIR2:S39656	ck: 2631	len: 252	finds: 1	i	lipoteichoic acid biosynthe
PIR2:S72921	ck: 2352	len: 86	finds: 1	i	hypothetical protein B2168-
PIR2:S70669	ck: 7905	len: 381	finds: 1	i	coenzyme F420-dependent NS,
PIR2:C70828	ck: 2874	len: 97	finds: 1	i	hypothetical protein RV0463

PIR2:D70510	ck: 2571	len: 300	finds: 1	! hypothetical protein RV0519c	PIR2:S73488	ck: 5866	len: 385	finds: 1	! hypothetical protein K05.or
PIR2:F70547	ck: 3126	len: 294	finds: 1	! hypothetical protein RV0547c	PIR2:A36149	ck: 5138	len: 237	finds: 1	! spiralin - Spiroplasma mell
PIR2:C70708	ck: 4949	len: 259	finds: 1	! hypothetical protein RV0776c	PIR2:H69019	ck: 3145	len: 157	finds: 1	! conserved hypothetical prot
PIR2:G70716	ck: 883	len: 282	finds: 1	! hypothetical protein RV0953c	PIR2:D69049	ck: 1234	len: 355	finds: 1	! conserved hypothetical prot
PIR2:D70897	ck: 9641	len: 103	finds: 1	! hypothetical protein RV1102c	PIR2:D69065	ck: 9991	len: 249	finds: 1	! conserved hypothetical prot
PIR2:F70958	ck: 4095	len: 212	finds: 1	! hypothetical protein RV1377c	PIR2:G64362	ck: 6767	len: 406	finds: 1	! 2-isopropylmalate synthase
PIR2:C70763	ck: 497	len: 148	finds: 1	! hypothetical protein RV1558	PIR2:G64456	ck: 7004	len: 167	finds: 1	! hypothetical protein homolo
PIR2:C70541	ck: 1907	len: 170	finds: 1	! hypothetical protein RV1577c	PIR2:D64423	ck: 9981	len: 329	finds: 1	! hypothetical protein Mj0988
PIR2:A70639	ck: 9160	len: 181	finds: 1	! hypothetical protein RV1957	PIR2:D64452	ck: 2920	len: 299	finds: 1	! hypothetical protein Mj1221
PIR2:B70758	ck: 1836	len: 317	finds: 1	! hypothetical protein RV1996	PIR2:F69386	ck: 9110	len: 437	finds: 1	! acyl-CoA dehydrogenase, sho
PIR2:A70941	ck: 7719	len: 239	finds: 1	! hypothetical protein RV2018	PIR2:G69440	ck: 1399	len: 136	finds: 1	! conserved hypothetical prot
PIR2:D70943	ck: 7125	len: 324	finds: 1	! hypothetical protein RV2037c	PIR2:A69470	ck: 6749	len: 61	finds: 1	! hypothetical protein AF1762
PIR2:H70863	ck: 4861	len: 419	finds: 1	! hypothetical protein RV2449c	PIR2:D69509	ck: 4730	len: 270	finds: 1	! hypothetical protein AF2077
PIR2:E70572	ck: 8454	len: 273	finds: 1	! hypothetical protein RV2622	PIR2:T08352	ck: 7002	len: 148	finds: 1	! hypothetical protein H1537
PIR2:F70572	ck: 9524	len: 297	finds: 1	! hypothetical protein RV2623	PIR2:S73242	ck: 4681	len: 291	finds: 1	! hypothetical protein 38 - r
PIR2:G70965	ck: 4149	len: 177	finds: 1	! hypothetical protein RV2651c	PIR2:S20471	ck: 4134	len: 86	finds: 1	! class V zygote-specific pro
PIR2:E70885	ck: 9649	len: 308	finds: 1	! hypothetical protein RV2859c	PIR2:S25969	ck: 7679	len: 69	finds: 1	! hypothetical protein 69 - 1
PIR2:F70669	ck: 5693	len: 418	finds: 1	! hypothetical protein RV2953	PIR2:T01551	ck: 1750	len: 289	finds: 1	! hypothetical protein A_1G00
PIR2:B70651	ck: 3649	len: 118	finds: 1	! hypothetical protein RV3073c	PIR2:T05405	ck: 4069	len: 222	finds: 1	! hypothetical protein F10M6
PIR2:D70645	ck: 2556	len: 332	finds: 1	! hypothetical protein RV3131	PIR2:T00559	ck: 9966	len: 296	finds: 1	! hypothetical protein F12I6
PIR2:F70951	ck: 8567	len: 355	finds: 1	! hypothetical protein RV3200c	PIR2:T01617	ck: 1414	len: 392	finds: 1	! hypothetical protein F19F24
PIR2:C70561	ck: 48	len: 350	finds: 1	! hypothetical protein RV3626c	PIR2:T06062	ck: 59	len: 169	finds: 1	! hypothetical protein F19H22
PIR2:H70850	ck: 8984	len: 254	finds: 1	! hypothetical protein RV3912	PIR2:T04272	ck: 5780	len: 349	finds: 1	! hypothetical protein F20B18
PIR2:A70565	ck: 5027	len: 226	finds: 1	! probable cutinase precursor	PIR2:T05639	ck: 5981	len: 445	finds: 1	! hypothetical protein F20D10
PIR2:D70696	ck: 8388	len: 326	finds: 1	! probable dtgp-glucose 4 - Myc	PIR2:T01144	ck: 6323	len: 272	finds: 1	! hypothetical protein F26B6
PIR2:B70986	ck: 9024	len: 203	finds: 1	! probable isomerase - Mycobact	PIR2:T01145	ck: 9711	len: 260	finds: 1	! hypothetical protein F26B6
PIR2:H70760	ck: 2319	len: 202	finds: 1	! probable lipoprotein signal pe	PIR2:T08549	ck: 4270	len: 404	finds: 1	! hypothetical protein F27B13
PIR2:B70763	ck: 4418	len: 397	finds: 1	! probable membrane protein - My	PIR2:T05096	ck: 9032	len: 171	finds: 1	! hypothetical protein F28W20
PIR2:A70614	ck: 5098	len: 287	finds: 1	! probable mmaA2 protein - Myco	PIR2:T02589	ck: 3673	len: 381	finds: 1	! hypothetical protein T16B24
PIR2:S13338	ck: 4079	len: 256	finds: 1	! hypothetical protein (insertio	PIR2:T02510	ck: 8930	len: 96	finds: 1	! hypothetical protein T19C21
PIR2:PC4110	ck: 2351	len: 137	finds: 1	! transcription regulator homolo	PIR2:T02122	ck: 8329	len: 248	finds: 1	! hypothetical protein T20B5
PIR2:PN0644	ck: 9954	len: 66	finds: 1	! hypothetical protein 66 - Stre	PIR2:T02548	ck: 5743	len: 217	finds: 1	! hypothetical protein T26B15
PIR2:S44233	ck: 6765	len: 358	finds: 1	! strp protein - Streptomyces g1	PIR2:T05859	ck: 7427	len: 208	finds: 1	! hypothetical protein T29A15
PIR2:S17674	ck: 4817	len: 291	finds: 1	! serine-type D-Ala-D-Ala carbox	PIR2:T01365	ck: 1733	len: 175	finds: 1	! hypothetical protein T29F13
PIR2:S17177	ck: 5969	len: 280	finds: 1	! rRNA methylase - Streptomyces	PIR2:T08993	ck: 2816	len: 364	finds: 1	! hypothetical protein T32A16
PIR2:C64251	ck: 2836	len: 385	finds: 1	! hypothetical protein homolog M	PIR2:T06094	ck: 7672	len: 178	finds: 1	! hypothetical protein T5017
PIR2:I64230	ck: 1048	len: 248	finds: 1	! sensory rhodopsin II transduce					

PIR2:T00988	ck: 916	len: 241	finds: 1	! hypothetical protein T9J2.22
PIR2:T08191	ck: 2415	len: 415	finds: 1	! probable translation elongation
PIR2:T09662	ck: 3889	len: 101	finds: 1	! multicatalytic endopeptidase
PIR2:S23774	ck: 5889	len: 402	finds: 1	! uricose phosphate/3-phosphoglyc
PIR2:S22500	ck: 9347	len: 256	finds: 1	! embryonic protein ecp31 - can
PIR2:T16984	ck: 4925	len: 165	finds: 1	! transcription factor homolog
PIR2:S47086	ck: 1305	len: 263	finds: 1	! p17a protein - rice
PIR2:S14884	ck: 7397	len: 146	finds: 1	! hypothetical protein 7 - yeast
PIR2:S43448	ck: 5386	len: 215	finds: 1	! FUN54 protein precursor - yeast
PIR2:S19425	ck: 1969	len: 317	finds: 1	! hypothetical protein YCR015c -
PIR2:S70131	ck: 1796	len: 365	finds: 1	! hypothetical protein YDR273w -
PIR2:S69699	ck: 3355	len: 374	finds: 1	! hypothetical protein YDR415c -
PIR2:S50587	ck: 5798	len: 128	finds: 1	! hypothetical protein YER084w -
PIR2:S64117	ck: 6897	len: 107	finds: 1	! hypothetical protein YGL109w -
PIR2:S46801	ck: 5569	len: 429	finds: 1	! hypothetical protein YKL027w -
PIR2:S64829	ck: 7470	len: 336	finds: 1	! hypothetical protein YLR007w -
PIR2:S68481	ck: 4630	len: 254	finds: 1	! hypothetical protein YLR165c -
PIR2:S69848	ck: 7663	len: 105	finds: 1	! hypothetical protein YML101c -
PIR2:S77699	ck: 7370	len: 238	finds: 1	! inner cell wall mannoprotein
PIR2:A40550	ck: 166	len: 441	finds: 1	! peroxisomal membrane protein
PIR2:S67699	ck: 4697	len: 193	finds: 1	! probable membrane protein YDL1
PIR2:S60916	ck: 7226	len: 139	finds: 1	! probable membrane protein YNL2
PIR2:S66918	ck: 8726	len: 157	finds: 1	! probable membrane protein YOR0
PIR2:JT0589	ck: 609	len: 331	finds: 1	! pectinesterase (EC 3.1.1.11) F
PIR2:S03833	ck: 3269	len: 319	finds: 1	! hypothetical protein 1 - chest
PIR2:A34051	ck: 7750	len: 201	finds: 1	! NADH dehydrogenase (ubiquinone
PIR2:S16556	ck: 4238	len: 289	finds: 1	! hypothetical protein 1 - Podos
PIR2:S16557	ck: 4187	len: 289	finds: 1	! hypothetical protein 2 - Podos
PIR2:JN0451	ck: 2342	len: 312	finds: 1	! phosphoribosylanthranilate isc
PIR2:B44418	ck: 3830	len: 412	finds: 1	! surface antigen - Trypanosoma
PIR2:T14622	ck: 9305	len: 425	finds: 1	! hypothetical protein - Trypano
PIR2:T15458	ck: 2648	len: 331	finds: 1	! hypothetical protein C08A9.7 -
PIR2:T15482	ck: 121	len: 217	finds: 1	! hypothetical protein C10A4.3 -
PIR2:T15182	ck: 9118	len: 150	finds: 1	! hypothetical protein C18E3.4 -
PIR2:T15715	ck: 4709	len: 280	finds: 1	! hypothetical protein C30G12.2 -
PIR2:T15729	ck: 770	len: 304	finds: 1	! hypothetical protein C31H1.6 -
PIR2:T16486	ck: 1289	len: 374	finds: 1	! hypothetical protein F56D3.1 -
PIR2:T16528	ck: 3866	len: 419	finds: 1	! hypothetical protein K02F3.
PIR2:S40941	ck: 8579	len: 219	finds: 1	! hypothetical protein ZK632.
PIR2:S44857	ck: 5825	len: 170	finds: 1	! PAR2.1 protein - Caenorhabd
PIR2:S44865	ck: 399	len: 336	finds: 1	! R0SD3.9 protein - Caenorhab
PIR2:S06613	ck: 4664	len: 196	finds: 1	! chorion protein sl9 - fruit
PIR2:S69241	ck: 7903	len: 298	finds: 1	! Dreg-5 protein - fruit fly
PIR2:A24254	ck: 68	len: 255	finds: 1	! ecdysone-induced protein -
PIR2:S78042	ck: 1000	len: 351	finds: 1	! Ig mu chain C region, membr
PIR2:S69288	ck: 5524	len: 102	finds: 1	! early lactation protein pre
PIR2:T14143	ck: 3550	len: 68	finds: 1	! ATPase subunit 8 - oranguta
PIR2:JC5761	ck: 447	len: 225	finds: 1	! cytokine-inducible SH2 prot
PIR2:S43928	ck: 3434	len: 310	finds: 1	! estradiol 17beta-dehydrogen
PIR2:JE0175	ck: 3945	len: 317	finds: 1	! frezzled protein-1b - human
PIR2:T77373	ck: 9167	len: 22	finds: 1	! gene N-ras protein - human
PIR2:T14788	ck: 6935	len: 112	finds: 1	! hypothetical protein DKF2p5
PIR2:T08722	ck: 6844	len: 381	finds: 1	! hypothetical protein DKF2p5
PIR2:T00702	ck: 4343	len: 196	finds: 1	! hypothetical protein F25965
PIR2:A44478	ck: 3379	len: 373	finds: 1	! probable cell growth or dif
PIR2:B44478	ck: 908	len: 337	finds: 1	! probable cell growth or dif
PIR2:A61188	ck: 916	len: 358	finds: 1	! probable transcription fact
PIR2:S63679	ck: 1615	len: 65	finds: 1	! signal transducer and activ
PIR2:A41795	ck: 8652	len: 391	finds: 1	! somatostatin receptor 1 - h
PIR2:SC5627	ck: 178	len: 225	finds: 1	! STAR induced STAR inhibitor
PIR2:S41063	ck: 8093	len: 183	finds: 1	! translocan-associated prote
PIR2:A36679	ck: 8392	len: 183	finds: 1	! signal sequence receptor be
PIR2:A56446	ck: 3609	len: 268	finds: 1	! Ig heavy chain V region (3H
PIR2:B61188	ck: 9720	len: 265	finds: 1	! SCL protein - mouse (fragme
PIR2:S63681	ck: 840	len: 64	finds: 1	! signal transducer and activ
PIR2:C41795	ck: 8110	len: 391	finds: 1	! somatostatin receptor 1 - m
PIR2:S66270	ck: 7215	len: 457	finds: 1	! kynurenine--oxoglutarate tr
PIR2:S41286	ck: 8455	len: 223	finds: 1	! latexin - rat
PIR2:A39297	ck: 9240	len: 391	finds: 1	! somatostatin receptor - rat
PIR2:SC4294	ck: 404	len: 274	finds: 1	! ammonia monooxygenase (EC 1
PIR2:C75042	ck: 9304	len: 308	finds: 1	! aspartate carboxyltransfer
PIR2:D71333	ck: 4876	len: 477	finds: 1	! conserved hypothetical prot
PIR2:G71287	ck: 2410	len: 365	finds: 1	! conserved hypothetical prot

```

PIR2:J00857 ck: 6140 len: 169 finds: 1 i hypothetical 18.1K protein - B
PIR2:T10288 ck: 562 len: 298 finds: 1 i hypothetical protein 19 - Orgy
PIR2:G72757 ck: 9964 len: 175 finds: 1 i hypothetical protein APE0052
PIR2:C72759 ck: 3453 len: 114 finds: 1 i hypothetical protein APE0064
PIR2:A72773 ck: 7998 len: 129 finds: 1 i hypothetical protein APE0172
PIR2:C72775 ck: 1302 len: 100 finds: 1 i hypothetical protein APE0189
PIR2:C72784 ck: 9325 len: 313 finds: 1 i hypothetical protein APE0258
PIR2:H72721 ck: 4962 len: 168 finds: 1 i hypothetical protein APE0313
PIR2:H72731 ck: 2984 len: 172 finds: 1 i hypothetical protein APE0393
PIR2:G72736 ck: 2336 len: 180 finds: 1 i hypothetical protein APE0429
PIR2:D72737 ck: 7680 len: 120 finds: 1 i hypothetical protein APE0434
PIR2:A72741 ck: 338 len: 160 finds: 1 i hypothetical protein APE0458
PIR2:G72640 ck: 9276 len: 147 finds: 1 i hypothetical protein APE0559
PIR2:A72651 ck: 639 len: 119 finds: 1 i hypothetical protein APE0638
PIR2:D72672 ck: 4161 len: 129 finds: 1 i hypothetical protein APE0802
PIR2:E72679 ck: 8681 len: 234 finds: 1 i hypothetical protein APE0857
PIR2:F72685 ck: 9005 len: 228 finds: 1 i hypothetical protein APE0902
PIR2:H72607 ck: 1372 len: 387 finds: 1 i hypothetical protein APE1326
PIR2:E72626 ck: 6841 len: 146 finds: 1 i hypothetical protein APE1469
PIR2:D72630 ck: 4651 len: 152 finds: 1 i hypothetical protein APE1499
PIR2:H72538 ck: 9362 len: 175 finds: 1 i hypothetical protein APE1597
PIR2:D72545 ck: 2785 len: 194 finds: 1 i hypothetical protein APE1648
PIR2:E72557 ck: 8347 len: 132 finds: 1 i hypothetical protein APE1745
PIR2:H72557 ck: 8142 len: 181 finds: 1 i hypothetical protein APE1748
PIR2:D72577 ck: 2318 len: 100 finds: 1 i hypothetical protein APE1900
PIR2:A72511 ck: 2835 len: 140 finds: 1 i hypothetical protein APE2063
PIR2:C72529 ck: 9995 len: 310 finds: 1 i hypothetical protein APE2207
PIR2:F72471 ck: 430 len: 133 finds: 1 i hypothetical protein APE2415
PIR2:C70405 ck: 5299 len: 499 finds: 1 i hypothetical protein aq_1220
PIR2:D70454 ck: 3320 len: 269 finds: 1 i hypothetical protein aq_1793
PIR2:A75196 ck: 70 len: 147 finds: 1 i hypothetical protein PAB0052
PIR2:E75204 ck: 8207 len: 179 finds: 1 i hypothetical protein PAB0107
PIR2:H75185 ck: 6767 len: 266 finds: 1 i hypothetical protein PAB2346
PIR2:D71228 ck: 1741 len: 135 finds: 1 i hypothetical protein PH0086
PIR2:D71110 ck: 6672 len: 102 finds: 1 i hypothetical protein PH0651
PIR2:E71120 ck: 8797 len: 104 finds: 1 i hypothetical protein PH0732
PIR2:H71124 ck: 5357 len: 272 finds: 1 i hypothetical protein PH0767

PIR2:F71127 ck: 8012 len: 183 finds: 1 i hypothetical protein PH0787
PIR2:F71039 ck: 7274 len: 447 finds: 1 i hypothetical protein PH1606
PIR2:F71201 ck: 6730 len: 137 finds: 1 i hypothetical protein PH1883
PIR2:D71216 ck: 2646 len: 225 finds: 1 i hypothetical protein PH1966
PIR2:G75035 ck: 3225 len: 331 finds: 1 i iron (iii) abc transporter,
PIR2:B75110 ck: 5732 len: 249 finds: 1 i lipote-protein ligase A re
PIR2:E49094 ck: 3704 len: 373 finds: 1 i methylmalonyl-CoA decarboxy
PIR2:T14222 ck: 7879 len: 344 finds: 1 i NADH dehydrogenase subunit
PIR2:S55472 ck: 8743 len: 299 finds: 1 i pectin lyase (EC 4.2.2.10)
PIR2:D72485 ck: 7054 len: 242 finds: 1 i probable high-affinity bran
PIR2:B71024 ck: 6905 len: 249 finds: 1 i probable lipote protein 11
PIR2:S72288 ck: 9584 len: 75 finds: 1 i ribosomal protein L23 - Pla
PIR2:T10349 ck: 889 len: 374 finds: 1 i very-late factor 1 protein
PIR2:S78187 ck: 3067 len: 197 finds: 1 i ymf39 protein - RecJinmona
PIR3:T11851 ck: 3584 len: 324 finds: 1 i ruvb protein - Thermus aqua
PIR4:JCS705 ck: 3534 len: 125 finds: 1 i alpha-amylase inhibitor 0.1

\\End of list
Databases searched:
NRBF, Release 62.0, Released on 30sep1999, Formatted on 26oct1999

Total finds: 561
Total lengths: 47,169,319
Total sequences: 142,080
CPU time: 05:21.97

i FINDPATTERNS on PIR: * allowing 0 mismatches
i 1 <(x) [1,200] (L,I,V,M,A,P)X(P,T,S) (L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) (C,T,S) (R,
CCOF2T ck: 2620 len: 92 i cytochrome c2 - Rhodocyclus tenuis
1 <(x) [1,200] (L,I,V,M,A,P)X(P,T,S) (L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
x[64] (P)X(P) (A)XX(V) (T) (K)XX(A)XX(V)X[12]
ADBSALAIQTRGCLACHINPEKRYVGRPAVAKKVGAGAEKLVAKVMAGGOGWMAQOLG
1: RDHUCB ck: 7216 len: 277 i carbonyl reductase (NADPH) (EC 1.1.1.184
<(x) [1,200] (L,I,V,M,A,P)X(P,T,S) (L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
x[1189] (P)X(S) (A)XX(V) (T) (K)XX(V)XX(S)X[72]
MSSGTHVALVTCGKNGGIGLAIVRDLGSLFGSDVLTNRDVTIRGAAVQOQAGAEGLSRFHOLD
1: DTKRGVQKKEGMSAAGVTKICVYVTSRTHAKISBORGDKILNACCPCGVKRI
DMAQPKATKSEEGAEETPVILALPPDAEGPHGQFVSEKRVQEM
RDALAE ck: 2829 len: 246 i acetoacetyl-CoA reductase (EC 1.1.1.36)
1 <(x) [1,200] (L,I,V,M,A,P)X(P,T,S) (L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
x[112] (L)X(S) (L)XX(V) (T) (K)XX(I)XX(A)X[118]
MTORLAAYTGGMGIGICORLAKDGRVAVAGCGPSPREKWLFOOKALGDFIASSEGNA
VNTVSPGIATDMKAIROVDLKRIVATIPVKRLGIPETIASICAMLSSEESGFSF
GADSLNGGLHMG

```







[illegible]





1	539654	ck: 5853	len: 255	1	hypothetical protein X - Pseudomonas aeruginosa	1	JC5439	ck: 934	len: 282	1	D-aspartate oxidase (EC 1.4.3.1) type II	1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{103}(A)x(S)(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{146}(L)x(T)(A)x(T)(S)(R)xx(A)xxx(T)x{136} MNPRLNGKGVAVLTGARGIGISITMLIAEOMVYVLANDBREGRARVAALGSEHVAVDVA DIRVNAISPGWIDREARERAAALITELDBDQHLVGVNGVEDVASYANVLSEDA GFVYQGEFLVNGMTRKRIIYD
1	H69885	ck: 6790	len: 242	1	3-oxoacyl-acyl-carrier protein reductase	1	S25422	ck: 6471	len: 72	1	NADH dehydrogenase (ubiquinone) (EC 1.6.1.1)	1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{109}(V)x(S)(P)x(X)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{146}(L)x(T)(A)x(T)(S)(R)xx(A)xxx(T)x{117} MNPRLNGKGVAVLTGARGIGISITMLIAEOMVYVLANDBREGRARVAALGSEHVAVDVA DIRVNAISPGWIDREARERAAALITELDBDQHLVGVNGVEDVASYANVLSEDA GFVYQGEFLVNGMTRKRIIYD
1	F69868	ck: 7200	len: 248	1	glucose-1-dehydrogenase homolog ykwo - Bacillus subtilis	1	S08622	ck: 675	len: 307	1	hydrogenase (EC 1.18.99.1) 3 chain 4 - E. coli	1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{120}(P)x(T)(A)x(X)(A)(S)(R)xx(A)xxx(L)x{171} MSVLPRLGALVLELGLSITRVARALNRGGVGLQETRDITKLLGROSVGPDAAGVNW ALCAAPATPIEMKFLPDLAEBOELQESPLVSGSPGWKWSISKOLVQ MFEVFFIPWGMETFTTGLLVLVAVLVGVVIALFENSARLRLDITPRITVAGAFALFVSLAA TTHGFEASAMVWHFVDVWVLFLETTIYMWGN
1	B55850	ck: 1047	len: 267	1	hypothetical protein (hshd 5' region) - Cl. botulinum	1	T11338	ck: 2220	len: 347	1	NADH dehydrogenase (ubiquinone) (EC 1.6.1.1)	1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{49}(P)x(S)(I)x(X)(A)(T)(R)xx(L)xxx(T)x{282} MNPRLNGKGVAVLTGARGIGISITMLIAEOMVYVLANDBREGRARVAALGSEHVAVDVA DIRVNAISPGWIDREARERAAALITELDBDQHLVGVNGVEDVASYANVLSEDA GFVYQGEFLVNGMTRKRIIYD
1	A56275	ck: 101	len: 387	1	1,3-propanediol dehydrogenase (EC 1.1.1.202)	1	T11159	ck: 979	len: 114	1	NADH dehydrogenase 3 - hardbacked tick	1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{73}(P)x(P)(L)xx(M)(T)(R)xx(L)xxx(S)x{25} MFEVFFIPWGMETFTTGLLVLVAVLVGVVIALFENSARLRLDITPRITVAGAFALFVSLAA TTHGFEASAMVWHFVDVWVLFLETTIYMWGN
1	S47829	ck: 4443	len: 339	1	glycerol-3-phosphate dehydrogenase (NAD+)	1	F64999	ck: 2909	len: 184	1	NADH dehydrogenase (ubiquinone) (EC 1.6.1.1)	1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{68}(L)x(T)(A)x(X)(A)(S)(R)xx(L)xxx(P)x{255} MNPRLNGKGVAVLTGARGIGISITMLIAEOMVYVLANDBREGRARVAALGSEHVAVDVA DIRVNAISPGWIDREARERAAALITELDBDQHLVGVNGVEDVASYANVLSEDA GFVYQGEFLVNGMTRKRIIYD
1	A44132	ck: 6742	len: 338	1	D-aspartate oxidase (EC 1.4.3.1) - bovine	1	S52970	ck: 6545	len: 167	1	NADH dehydrogenase (ubiquinone) (EC 1.6.1.1)	1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{15}(L)x(X)(L)xx(L)(S)(R)xx(L)xxx(L)x{307} MNPRLNGKGVAVLTGARGIGISITMLIAEOMVYVLANDBREGRARVAALGSEHVAVDVA DIRVNAISPGWIDREARERAAALITELDBDQHLVGVNGVEDVASYANVLSEDA GFVYQGEFLVNGMTRKRIIYD
1	JC5438	ck: 7167	len: 341	1	D-aspartate oxidase (EC 1.4.3.1) type III	1	S78138	ck: 8115	len: 267	1	cytochrome-c oxidase (EC 1.9.3.1) chain	1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{155}(L)x(T)(A)x(T)(S)(R)xx(L)xxx(T)x{196} MNPRLNGKGVAVLTGARGIGISITMLIAEOMVYVLANDBREGRARVAALGSEHVAVDVA DIRVNAISPGWIDREARERAAALITELDBDQHLVGVNGVEDVASYANVLSEDA GFVYQGEFLVNGMTRKRIIYD

[illegible]

1	IS1665	ck: 4724	len: 360	1	arginase 3 - African clawed frog
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{90}(P)x(T)(V)xx(A)(C)(K)xx(A)xxx(G)x{254} MSIRSFVFLKKNVNIKLQKCSHSAVIGAPSKQKRRGVEHGPALRSAGLIDRLSNLGN PPIPGFSWAKPCLSKSDIVYIGLRDLDPAEOPILKKNVNIKSYSMRHADCGIRKVM EXTFDOLGRDRPIHLSFDIDADDPALAPATGPVIGGLYRGGVITTEIHNTGMLSDLDVEVNPVLATTESEVKATANLAVDVI			
1	JS0609	ck: 8425	len: 140	1	biastictidin-S deaminase (EC 3.5.4.23) - Bac
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{95}(V)x(S)(P)xx(M)(C)(R)xx(I)xxx(A)x{129} MKTFNISQGDDELVEVATEKIMLEEDKHHGALIRKTEIISAVHIEAYIGRVVCAEALAIIG			
1	S37304	ck: 2135	len: 335	1	spal protein - Salmonella typhimurium (frag
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{24}(V)xx(P)(P)xx(A)(S)(R)xx(V)xxx(L)x{295} DPLGKIVERTFEVAPISERVIDAPPSVSRVGRREPLITGVRAIDGLITCGVGQRMGIFASAG YPAVSFNDLPRLLEPGATSEGSTIAFTVVLSEEDADPADIRSIIDGHLYS RKLAGQGHYPAIDVLSKYSRVFGQVTTPTHAEOASAVRKLMTLLELQFLIDGEXRGENINDRAMQWRDLSKAMLCQPVAYSSF			
1	S01397	ck: 4392	len: 127	1	H+-transporting ATP synthase (EC 3.6.1.34)
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{68}(V)x(T)(L)xx(L)(S)(R)xx(L)xxx(A)x{43} MGNLQAMFMRQVRITLILATYTLTGFGFTPKVTVFSLITGTSISLWMVNLTKIEFGQAVV			
1	S22348	ck: 1686	len: 168	1	H+-transporting ATP synthase (EC 3.6.1.34)
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{141}(V)x(T)(A)xx(A)(T)(R)xx(L)xxx(I)x{11} MLPAAHLRPRGLRKLRAKRAVAAEAAAAPAAAGSPNOMSFIFASTOYFFNGANVRQVDVPTLTGA			
1	D72121	ck: 2102	len: 209	1	ATP synthase chain D - Chlamydia pneumoniae
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{119}(L)x(T)(M)xx(A)(S)(R)xx(V)xxx(V)x{74} MSVOYKLITNSFRLEROKLARLQIYTLPLIKLKALQAEVONAVKDAEEDKDYQAYERITAFAB			
1	VFLSDRSITDVGQVMAKKIIELRKARDECV	1:			
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{5}(V)xx(S)(A)xx(A)(T)(R)xx(V)xxx(S)x{160} MASTVSSAAVATRSNVQAQSMVAFETLKSASAPVTKKNNNDIISLMSNGSRVACMQVMPPIN			
1	S31498	ck: 4433	len: 181	1	ribulose-bisphosphate carboxylase (EC 4.1.1
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{5}(V)xx(S)(A)xx(A)(T)(R)xx(V)xxx(S)x{160} MASTVSSAAVATRSNVQAQSMVAFETLKSASAPVTKKNNNDIISLMSNGSRVACMQVMPPIN			
1	A71295	ck: 255	len: 357	1	probable alanine racemase (alr) - syphilis
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{184}(A)x(S)(A)xx(L)(C)(I)xx(A)xxx(M)x{157} MSRTRARVCLPVRKAAYAGGACDVAAQALSCGVHFAVACQOASQRAAGVAPILCLSTPAEF SIPLVHAANSALLCHPRAHFDVMPVPGGLAAGVAPESVHPARVSLPVELVTV			
1	RAIKRIPAGAVYSYORLWRAHTEHVGLPIGADGVMAALSPGLQVIGKWPVYVGAICMDQCVVDLGPLRVTVGSDRTLFGSPD	1:			
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{30}(P)x(S)(L)xx(L)(T)(K)xx(L)xxx(A)x{122} MSNYRDEFTIYVPGIAYVLADGSVLTIDRESINSLKPLVLAQMAALISGAADPRVTKVVRIGGA AAEALSMGIVASVAPADDEFQVNDKISRLSLAGKALAIATKKNINATITLTELAPT			
1	G70553	ck: 8877	len: 268	1	probable echA10 protein - Mycobacterium tub
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{30}(P)x(S)(L)xx(L)(T)(K)xx(L)xxx(A)x{122} MSNYRDEFTIYVPGIAYVLADGSVLTIDRESINSLKPLVLAQMAALISGAADPRVTKVVRIGGA AAEALSMGIVASVAPADDEFQVNDKISRLSLAGKALAIATKKNINATITLTELAPT			
1	H69457	ck: 4102	len: 322	1	ornithine cyclodeaminase (arcB) homo1og
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{178}(V)x(P)(A)xx(A)(S)(R)xx(V)xxx(T)x{128} MTELLITQEEVESLISMDQANNAVEAEARLYALCKAQMPPRVYLLEFEKGDLRAMPALHMGYAG SVQPAEASRCQDVLTPTPSRKPVRKAEWVEEGTHIVAGIDAGDGPQEDLVELLK AKIVVDLEQAKHGGEINVAVSKGIVIEVDVHATIGEIVAGLKDGRSEDEIIFDSTGLAIDQVAAVAVYENALSKRVGSKIK			
1	S75089	ck: 8968	len: 338	1	UDP-glucose 4-epimerase - Synechocystis
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{141}(P)x(S)(P)xx(A)(S)(K)xx(V)xxx(L)x{181} MATOCTIIVTGAGAGTSGHGVLAQAGFDVLTIDNLSYGHKELVQPLGVELVYVGHGDRQKL GADPGRLGEDNPFETHILPLALTLAKORPOLSVRGTDVDTLDGSTRALNDYTHVCD LAIAHYLGQYLLGSGESNIFNLGNGSFVRQVIEAKAVTGLDIPYQLCPRPRDAPILVGSSAKAREILGMNPQYRDLHTII			
1	H69105	ck: 3451	len: 336	1	dTPD-glucose 4,6-dehydratase - Methanoba
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{142}(P)x(S)(P)xx(A)(S)(K)xx(A)xxx(V)x{178} MEKILVYGAGPAGISGNFIRYMLQHPHYHILMDALTYCGNLEMLRGVEDPRRYTFVRSITDR YGPYOFPEKLIPLMITNALKENKPLPVYGDGMVMDWIHVHDRCRAVDLYLHGRVG EYVNIIGNSERNEIIVELIYRELKDESDLREVEDRPHDRYAIADASKIRNELGMPKPLVSFEEDIRITRMVYIDNRDWMENIK			
1	D69290	ck: 128	len: 332	1	dTPD-glucose 4,6-dehydratase (rfdB) homo
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{142}(P)x(S)(P)xx(A)(S)(K)xx(A)xxx(V)x{174} MRLVYTGAGTSGHGVLAQAGFDVLTIDNLSYGHKELVQPLGVELVYVGHGDRQKL YGPYOFPEKLIPLMITNALKENKPLPVYGDGMVMDWIHVHDRCRAVDLYLHGRVG EYVNIIGNSERNEIIVELIYRELKDESDLREVEDRPHDRYAIADASKIRNELGMPKPLVSFEEDIRITRMVYIDNRDWMENIK			
1	S47045	ck: 5086	len: 346	1	dTPD-glucose 4,6-dehydratase (EC 4.2.1.46
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{160}(P)x(S)(P)xx(A)(S)(K)xx(A)xxx(V)x{170} MOTBEKKNILVTGGAGTSGHGVLAQAGFDVLTIDNLSYGHKELVQPLGVELVYVGHGDRQKL VEDHARLYQVYTGAGTSGHGVLAQAGFDVLTIDNLSYGHKELVQPLGVELVYVGHGDRQKL YGPYOFPEKLIPLMITNALKENKPLPVYGDGMVMDWIHVHDRCRAVDLYLHGRVG EYVNIIGNSERNEIIVELIYRELKDESDLREVEDRPHDRYAIADASKIRNELGMPKPLVSFEEDIRITRMVYIDNRDWMENIK			
1	E70566	ck: 5640	len: 331	1	probable dTPD-GLUCOSE 4 - Mycobacterium
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{142}(P)x(S)(P)xx(A)(T)(K)xx(A)xxx(V)x{173} MRLVYTGAGTSGHGVLAQAGFDVLTIDNLSYGHKELVQPLGVELVYVGHGDRQKL YGPYOFPEKLIPLMITNALKENKPLPVYGDGMVMDWIHVHDRCRAVDLYLHGRVG EYVNIIGNSERNEIIVELIYRELKDESDLREVEDRPHDRYAIADASKIRNELGMPKPLVSFEEDIRITRMVYIDNRDWMENIK			
1	G70415	ck: 2675	len: 321	1	nucleotide sugar epimerase - Aquifex aeo
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{146}(P)x(S)(P)xx(A)(S)(K)xx(A)xxx(V)x{159} MRLVYTGAGTSGHGVLAQAGFDVLTIDNLSYGHKELVQPLGVELVYVGHGDRQKL YGPYOFPEKLIPLMITNALKENKPLPVYGDGMVMDWIHVHDRCRAVDLYLHGRVG EYVNIIGNSERNEIIVELIYRELKDESDLREVEDRPHDRYAIADASKIRNELGMPKPLVSFEEDIRITRMVYIDNRDWMENIK			
1	E69750	ck: 3388	len: 249	1	glucosamine-6-phosphate isomerase homo1o
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{184}(I)x(T)(I)xx(F)(S)(K)xx(V)xxx(S)x{49} MRLVYTGAGTSGHGVLAQAGFDVLTIDNLSYGHKELVQPLGVELVYVGHGDRQKL YGPYOFPEKLIPLMITNALKENKPLPVYGDGMVMDWIHVHDRCRAVDLYLHGRVG EYVNIIGNSERNEIIVELIYRELKDESDLREVEDRPHDRYAIADASKIRNELGMPKPLVSFEEDIRITRMVYIDNRDWMENIK			

D69094 ck: 5550 len: 248 1 phosphoribosylaminoimidazole succinocarboxan  
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x(122)(L)x(T)(I)x(W)(T)(K)x(A)xxx(T)x(277)  
 x(122)(L)x(T)(I)x(W)(T)(K)x(A)xxx(T)x(277)  
 MDVHIDGFLGSKAKADVLITDDPEIYARFRDITAGDEKDTLEMGYNSVIAKFEVLEEA  
 1: FKLEGGSSGRIRLGDGVSPTICRLMDMETGEPLDKDIFRGEEGVAGVARRVAR  
 MILDDEDIERMVEL  
 A19940 ck: 1681 len: 40 1 antithrombin III - baboon (fragment)  
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x(12)(V)x(T)(A)xx(W)(T)(K)xx(A)xxx(T)x(12)  
 DDKDNIFLSPLSVSTAFAMTKGACNDITKOLMEVEFKEDT  
 JX0206 ck: 7697 len: 207 1 chymotrypsin inhibitor (Kunitz) WCI-3 precu  
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x(69)(V)x(S)(P)xx(V)(S)(K)xx(P)xxx(S)x(122)  
 MKSTIFLAFLLSAITISHLPSSINDDDLVDAEGVLVNGGTYYLLPHMAHGSGIETAKTGNBPCH  
 1: RGNRRLVYTEENPLELVLLAKSETASSH  
 JS0650 ck: 8167 len: 207 1 chymotrypsin inhibitor (Kunitz) WCI-2 precu  
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x(69)(V)x(S)(P)xx(V)(S)(K)xx(P)xxx(S)x(122)  
 MKSTIFLAFLLSAITISHLPSSINDDDLVDAEGVLVNGGTYYLLPHMAHGSGIETAKTGNBPCH  
 1: RGNRRLVYTEENPLELVLLAKSETASSH  
 JX0246 ck: 1378 len: 181 1 serine proteinase inhibitor A precursor - a  
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x(14)(L)x(S)(L)xx(L)(C)(H)xx(P)xxx(S)x(115)  
 MAASNALITSGVLLISLAVLCHGDPVVDSDGDAVOLNLSGNYPLTYTQSAALGFRGSLTLHKDA  
 1: TAPQ  
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x(14)(L)x(S)(L)xx(L)(C)(H)xx(P)xxx(S)x(115)  
 MAASNALITSGVLLISLAVLCHGDPVVDSDGDAVOLNLSGNYPLTYTQSAALGFRGSLTLHKDA  
 1: JCS447 ck: 4632 len: 210 1 serine proteinase inhibitor A precursor - a  
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x(14)(L)x(S)(L)xx(L)(C)(H)xx(P)xxx(S)x(115)  
 MAASNALITSGVLLISLAVLCHGDPVVDSDGDAVOLNLSGNYPLTYTQSAALGFRGSLTLHKDA  
 1: FVNGRTLLIGIGEHFTVRFQKFDALAKMTAFO  
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x(14)(L)x(S)(L)xx(L)(C)(H)xx(P)xxx(S)x(115)  
 MAASNALITSGVLLISLAVLCHGDPVVDSDGDAVOLNLSGNYPLTYTQSAALGFRGSLTLHKDA  
 1: T06517 ck: 7537 len: 151 1 alpha-amylase inhibitor 1mal precursor, mot  
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x(120)(A)x(S)(V)xx(V)(C)(K)xx(L)xxx(S)x(115)  
 MMKTYFVGLVEMLVATMAVEYGARSHNSGSMWCDPATGIVASALTGCRAMVTLQCVGSVPE  
 1: S16920 ck: 2886 len: 121 1 alpha-amylase inhibitor - wheat  
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x(90)(A)x(S)(V)xx(V)(C)(K)xx(L)xxx(S)x(115)  
 SGPMSWCDPATGIVASALTGCRAMVTLQCVGSVPEAVLDDCCOQOLADINNEMWCGDLSMLR  
 1: A60195 ck: 2543 len: 278 1 transforming protein dbl - mouse (fragment)  
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x(146)(L)x(T)(I)x(W)(T)(K)x(A)xxx(T)x(277)  
 x(146)(L)x(T)(I)x(W)(T)(K)x(A)xxx(T)x(277)  
 DOSPKLNSGIDITKNTLVNLTOTERAVYRELFTVLGVRSMNDPQAFDMLPLNKRKDVLEFN  
 VWLGHKRGATKMDPFAFKPMQNHLLYKRAVWFCRRESECGADRPSYDFKHC  
 LKMDVQITSEHVKGNDKKEFIMTSEKEELIYQAPNVYKMLMK  
 1: A40316 ck: 7358 len: 415 1 cyclin B - fission yeast (Schizosaccharomy  
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x(122)(L)x(T)(I)x(W)(T)(K)x(A)xxx(T)x(277)  
 x(122)(L)x(T)(I)x(W)(T)(K)x(A)xxx(T)x(277)  
 MDVHIDGFLGSKAKADVLITDDPEIYARFRDITAGDEKDTLEMGYNSVIAKFEVLEEA  
 1: FKLEGGSSGRIRLGDGVSPTICRLMDMETGEPLDKDIFRGEEGVAGVARRVAR  
 MILDDEDIERMVEL  
 A57234 ck: 2377 len: 348 1 lin-44 protein precursor - Caenorhabdit  
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x(122)(L)x(T)(I)x(W)(T)(K)x(A)xxx(T)x(277)  
 x(122)(L)x(T)(I)x(W)(T)(K)x(A)xxx(T)x(277)  
 MDVHIDGFLGSKAKADVLITDDPEIYARFRDITAGDEKDTLEMGYNSVIAKFEVLEEA  
 1: FKLEGGSSGRIRLGDGVSPTICRLMDMETGEPLDKDIFRGEEGVAGVARRVAR  
 MILDDEDIERMVEL  
 J04152 ck: 4630 len: 354 1 wnt-11 protein precursor - chicken  
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x(115)(A)x(S)(A)xx(I)(S)(H)xx(A)xxx(T)x(223)  
 MKPSQFLAFLLSLILDTGICIGYIKMIALSKTPSSIALNQTQCHQLEGLVYVQVLCRSNL  
 1: MKKSGSQANKLMLHNSVGRQVLAKELEMKCHGVSCTIKTCMKGOELRDI  
 ALDKKRYLSATKVVHRPMGTRKYLVRKIDIRPVKTELIVLQSSPDECMKNERVSGHGTODRCNKTNSGSDCLMCCGNGY  
 1: S34378 ck: 1689 len: 354 1 wnt-11 protein - mouse  
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x(115)(A)x(S)(A)xx(I)(S)(H)xx(A)xxx(T)x(223)  
 MKPSQFLAFLLSLILDTGICIGYIKMIALSKTPSSIALNQTQCHQLEGLVYVQVLCRSNL  
 1: MKKSGSQANKLMLHNSVGRQVLAKELEMKCHGVSCTIKTCMKGOELRDI  
 ALDKKRYLSATKVVHRPMGTRKYLVRKIDIRPVKTELIVLQSSPDECMKNERVSGHGTODRCNKTNSGSDCLMCCGNGY  
 1: JH0687 ck: 9563 len: 398 1 bone morphogenetic protein 21 precursor  
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x(178)(V)x(P)(A)xx(A)(S)(R)xx(V)xxx(L)x(204)  
 MVAIGHSLDLDTFOTLISLSCGCTGLIPBEGKRYESGSRSSQOQVLDQFELRLNMGFLKR  
 1: IYKPAASASGPVRLDLRLIYHNHESKESFDTVPAIARWIAHROKNOHFEVLEY  
 HLDNDNVPRHVRISRLTLDKGMPIRPLVTFSHDGGALHROKROARHOKRRLKSSCRHPLVYDFSDVGMNDIYA  
 1: S21299 ck: 8703 len: 81 1 protein DaF8 - eastern green mamba  
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x(116)(L)x(T)(I)x(W)(T)(K)x(A)xxx(T)x(277)  
 x(116)(L)x(T)(I)x(W)(T)(K)x(A)xxx(T)x(277)  
 MDVHIDGFLGSKAKADVLITDDPEIYARFRDITAGDEKDTLEMGYNSVIAKFEVLEEA  
 1: FKLEGGSSGRIRLGDGVSPTICRLMDMETGEPLDKDIFRGEEGVAGVARRVAR  
 MILDDEDIERMVEL  
 A37045 ck: 6756 len: 189 1 thermostable direct hemolysin homolog TR  
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x(112)(L)x(S)(I)x(V)(S)(K)xx(A)xxx(T)x(161)  
 x(112)(L)x(S)(I)x(V)(S)(K)xx(A)xxx(T)x(161)  
 MKLTYFAFSLILAVSIFSIKSPALDLPISIPPSGSDLELVVANNITIKTESPVAVIADIW  
 1: A48942 ck: 5895 len: 189 1 thermostable direct hemolysin-related he  
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x(112)(L)x(S)(I)x(V)(S)(K)xx(A)xxx(T)x(161)  
 x(112)(L)x(S)(I)x(V)(S)(K)xx(A)xxx(T)x(161)  
 MKLTYFAFSLILAVSIFSIKSPALDLPISIPPSGSDLELVVANNITIKTESPVAVIADIW  
 1: NKLSFGCKSQI

1	J00472	ck: 2831	len: 135	1	T-cell receptor beta chain (BTB4) - bovine	
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C				
		x(24)(L)x(S)(P)xx(V)(S)(K)xx(A)xxx(I)x(95)				
1:		MISOHRDOSTLGAKRGALGLALSLSSPAVSKGASVTEICRALDFQATTVFWYRQFPRGLMLM				
1	J00473	ck: 3911	len: 136	1	T-cell receptor beta chain (BTB1) - bovine	
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C				
		x(23)(L)x(S)(P)xx(V)(S)(K)xx(A)xxx(I)x(97)				
1:		MISHNRDOSTLGATRKALGLALSLSSPAVSKGASVTEICRALDFQASSMFWYRQFPRGLVLM				
1	B53250	ck: 6952	len: 353	1	class I histocompatibility antigen BOLA-AW1	
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C				
		x(149)(V)x(T)(M)xx(I)(S)(K)xx(M)xxx(G)x(188)				
1:		LLSSGLVLETRGASHMRSTAVSRPRGGEVLETVGDDTQFVRDSDAPRPRRPRRMY				
1	CYEMLRRLYENKDTLLRADPPRAHVTTRHPISGREYTLKCMALGFYPERISLTWCR					
		NGEDQTDQDELVEYTRPSGDNFQKMAALLVPSGEOKYTCOVQHEGLQEPFLTKWEPQPSFLTGLIIVGLVLTGAVAGVVICM				
1	I50609	ck: 8972	len: 207	1	T-cell surface glycoprotein CD8 beta chain	
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C				
		x(140)(M)x(T)(L)xx(L)(T)(K)xx(M)xxx(P)x(51)				
1:		MARPMALMTLCLOLPFCFTNLSSQFPGYILTKTNSTTEIVCPMKGEHTGVYWMNOGRHFEPL				
1	GALLLLSLPTIRREYRLRLRLVRAHRR					
1	I39516	ck: 7310	len: 257	1	ribosomal protein S3 - Acholeplasma axanthu	
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C				
		x(92)(I)x(S)(L)xx(L)(T)(K)xx(I)xxx(G)x(149)				
1:		MGQKVNPIGFRVGIIRMDSDKWYADKKIYPALVLEKEDAVIRKFNKYNNAASHVIEIRLKLKAYK				
1	PLHTLRADVEYATAEQTYGILGIVMVIHGBILGDSHEELKROSSASSNHG					
		GGRRPSRKGRPSRSDAATBEGN				
1	S77489	ck: 8926	len: 133	1	ribosomal protein S8 - Synechocystis sp. (S	
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C				
		x(74)(I)x(T)(L)xx(V)(S)(K)xx(L)xxx(S)x(43)				
1:		MASDTDISMLTRIRNACAVRHSTITQVPTKMTLSIAKYLKSEGFIEDYSETGEGINKMLVLTLYK				
1	T01947	ck: 4063	len: 252	1	hypothetical protein Fl104.1 - Arabidopsis	
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C				
		x(77)(I)x(T)(P)xx(V)(S)(R)xx(A)xxx(A)x(229)				
1:		MPNIRALIGTPEASRPDAIRAPAFEFSSVITVIRFGQSGNAVGLTGDGPATPAGLVAAASLSHA				
1	ALGILVAGANTLGGADGASMKNAVSFGFAYVSMITNINWYWGPFICAAIAIY					
		YDTLFIQSNCHPEPLPSNDF				
1	A39234	ck: 9791	len: 371	1	opsin - bluebottle fly (Calliphora vicina)	
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C				
		x(117)(L)x(P)(V)xx(W)(S)(R)xx(P)xxx(L)x(176)				
1:		MERYSTLIGSFALNNGSVTDKVPDMAHLVAPYWNQFPAEPKAKFLAAYVWLATISWGN				
1	WTLLAPVFGMSRYVPEGNLSCGIDYLERQWNPNSYIIFISIVYVLPFLICYW					
		FILAAVSAHEKARBOAKKMNYSLSRSESDADKSEGLAKALVLTISLWMAVTPYIITLGLFKYEGLTPLNTIWGACFAKSAAC				
1	S40691	ck: 7181	len: 374	1	opsin rh1 - fruit fly (Drosophila pseudoobs	
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C				
		x(118)(L)x(P)(V)xx(W)(S)(R)xx(P)xxx(L)x(176)				
1:		MDSFAVATQIGRQFAAPSGSVYDKVTPPMALHISPYNDQDFAPAMPYAKILITAMITIGMSHC				
1	TIWCCILAPVFGMSRYVPEGNLSCGIDYLERQWNPNSYIIFISIVYVLPFLICY					
		SYFTILAAVSAHEKARBOAKKMNYSLSRSESDADKSEGLAKALVLTISLWMAVTPYIITLGLFKYEGLTPLNTIWGACF				
		JC4304	ck: 2863	len: 355	1	orphan G-protein-coupled receptor - huma
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W				
		x(80)(V)x(T)(L)xx(W)(T)(H)xx(I)xxx(G)x(259)				
1:		MDQEPESVLENFEYDGLADACVIGLVGFVLSIFSVFALIVGLNLLVVALNNSKRPK				
		DYDEVLOEIMPVLRNVEINLEGLFLPLILMSCYFRRIOTLFSCNKKRAKIKLI				
		LLVAVVIFLFWTPYNNMIFETLKLIDEFPSCDKMRDLALSVETVAFSHCCLNPLIAFAQEKFRRYLHLYGKCLAVLCGR				
1	I58186	ck: 4478	len: 354	1	hypothetical G-protein coupled receptor	
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W				
		x(81)(V)x(T)(L)xx(W)(T)(H)xx(I)xxx(G)x(257)				
1:		MPTSPPELENEVEDSDSADACVIGDIVAFGIFISIFSLVFTGLVGNLLVVALNNSKRS				
		GDYEVLOEIMPVLRNVEINLEGLFLPLILMSCYFRRIOTLFSCNKKRAKIKLI				
		ILLVAVVIFLFWTPYNNMIFETLKLIDEFPSCDKMRDLALSVETVAFSHCCLNPLIAFAQEKFRRYLHLYGKCLAVLCG				
1	A29667	ck: 8980	len: 60	1	pulmonary surfactant protein B - bovine	
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W				
		x(27)(A)x(T)(V)xx(V)(C)(H)xx(P)xxx(G)x(17)				
1:		FPPIPTCMLTRTLIRKIKQAVIPKRGVLAIVAOVCHVPLVGLIIOQLVIEYSVILXTD				
1	S22494	ck: 429	len: 277	1	rRNA N-glycosidase (EC 3.2.2.22) alpha-1	
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W				
		x(57)(L)x(S)(A)xx(A)(S)(R)xx(L)xxx(S)x(204)				
1:		MKRFTVILIALIPAASTVADVAFSSSGSSYSKFIDGLKRALPSNGVYNNITLLSSASG				
		AAASRRRYTEGOIIEISKNQVPSLATISLEMSALSKOIQLOATNGTKRTPV				
		ITDDKQRAVEITVTSKVYTKNIOLLNKNQVNAFDDVSAKH				
1	S23519	ck: 9157	len: 278	1	beta-luifin - smooth loofah	
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W				
		x(59)(L)x(S)(A)xx(A)(S)(R)xx(L)xxx(S)x(203)				
1:		MNRFTSLILIALFVAGVAVSSLSGADSKSKFTALRALPSKREKVSINPLLPSPA				
		TTAASRERYTGOIIEIRIPKNEVPSPALSLSEMSALSKOIQLOATNGNAFRTF				
		VVIIDNKGORVEIKDVNSKVYVNNIKLLNKNIAAFDDGLPTKH				
1	JN0108	ck: 7681	len: 250	1	luifin-b - smooth loofah	
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W				
		x(38)(L)x(S)(A)xx(A)(S)(R)xx(L)xxx(S)x(196)				
1:		ANYFSISGDSKSYSKFTALRAKALPSKREKVSINPLLPSPASGASRYITLMQSNYDAKATIM				
		NEVSPSPALSLSEMSALSKOIQLOATNGNAFRTFVVIIDNKGORVEITNLASKV				
		QINDVNSKILLNKNIA				
1	S65052	ck: 5285	len: 214	1	piestil-specific protein stsl4 precursor	
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W				
		x(43)(L)x(T)(A)xx(L)(T)(H)xx(A)xxx(P)x(155)				
1:		MFVLSITMAALVILYIYIDDEKREKLVKRNKTNLDFQFLLITASSLSHISAGVPPPP				
		KKSIEIGCAORCTEGPATLVCFNPNQVNGEKY				
1	JC4822	ck: 8148	len: 77	1	acyl carrier protein - Bacillus subtilis	
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W				
		x(A)x(T)(L)xx(V)(T)(K)xx(V)xxx(G)x(60)				
1:		MADTLERVTKIIVDRGLGVDEADVLTASFREDIGADSLDVLELMELEDFDEISDEDAE				
1	B65003	ck: 7651	len: 228	1	histidine transport system permease prot	
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W				

1:	KATOLAGKSTWEPFAIVCGIYLVFTTNSGVLLFLERRYSVGVKRAADL	x(125)(A)x(T)(A)xx(F)(T)(R)xx(V)xxx(1)x(87)	MLYGSVGLIGALATTELAISVLAIVIIIGLIGAGKLSQNSLGLIEGTTILIRGVPDLVMI
B36263	ck: 5147 len: 306	1 oligopeptide transport system permease prot	
1		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C	
1:		x(185)(I)x(S)(I)xx(I)(T)(R)xx(M)xxx(L)x(105)	
MYALSIAYASIRARITRSGSMIEVLHNSFTIRARAKGPMRIILRHAKPALLPV		MLKFTLRCLLEAPILFILLITISFEMRLALGSPTEGRTLPPEVANTEARXHLNDPIMTQYESY	
SYMGPAFVGIITGSVIEITIGLPGIGLFGALNDYSLVLTILVGLTILFENALIVDAVIDPKIRY			
C71163	ck: 4618 len: 304	1 probable oligopeptide transport system perm	
1		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C	
1:		x(55)(A)x(T)(L)xx(M)(T)(R)xx(I)xxx(T)x(233)	
MEYIKIAFNNKFKRGEGMLIAFIIFGLIPFPASDGLYEQVGIKIRIASKTLPMTREX			
ARSVRAQTLISLNNREYHNSKIMGVGDIPIYEDILPMISYIFMGIIQVSGAIL			
ASATLDFIGLGFITMVSGLVILQKAIMNALOFGMMWFIFIPGLFIILLITSLPFINLGIEVFNPRLGV			
S64728	ck: 6546 len: 482	1 protein secretion protein xcp - Pseudomona	
1		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C	
1:		x(151)(A)x(S)(A)xx(V)(S)(R)xx(V)xxx(L)x(315)	
MLPYRLARAGLAMPAGOGQWOLMRDQSEVLOELIRHAGPSALATIEPAPFEDQLAQYOGG			
GRITLRAAGREVDVRSITLPGIHGERVYRVLDKQSLALDNLGMPAAVADHLRS			
CLARPGVILSTGPTGSKTTLVLASLNDGSKNLTIVEDPEVAIGOTAINPRAGLTFASGLAIIRODPVIMGEIRDOH			
S52163	ck: 8969 len: 331	1 sucrose specific repressor - Escherichia co	
1		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C	
1:		x(24)(A)x(S)(V)xx(A)(T)(R)xx(V)xxx(1)x(291)	
MASLKDVARLAGVSMATVSRVHMAESVAPATRDVLOAIGTLVYDLSARKKRAQGRKPTLAV			
ROPLCEMLPESALATGRRGOFEOAMRBDGRLAEGFHNATGDHDTLASLNL			
AHRSKSGPDDVILICGNDBAAVAVOVLAKGVRIPOVAVMGEDNLVGHLEPLPLTITQDPHDIIGREAHLLIIEGREGVTRI			
D70044	ck: 4428 len: 243	1 transcription regulator GntR family homolog	
1		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C	
1:		x(124)(P)xx(S)(L)xx(I)(T)(R)xx(L)xxx(I)x(103)	
MNIKQSPILFYXOIMGKLTOKNGELOPDKPLPSREHYAEGEISMTVRQALSTLVNEGILYR			
ISRAKQLEPSAATTEBANILGIQKAPVLLIKRTIYLDNGIAFEHASKVYRGDRY			
S48826	ck: 9647 len: 215	1 MADH dehydrogenase (ubiquinone) (EC 1.6.5.3	
1		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C	
1:		x(36)(A)x(T)(P)xx(V)(T)(R)xx(P)xxx(S)x(163)	
NAMITRNATATLPLVLOSHRAAASHLHTSLPLSPATTPISITRPSPPTSSAPPGLSKTAEEVI			
VDIYVGCPTAEALTYGLLOLQKIRNRKDLHMMNK			
S75086	ck: 4153 len: 342	1 iron-stress chlorophyll-binding protein - S	
1		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C	
1:		x(43)(A)x(T)(L)xx(I)(S)(R)xx(P)xxx(M)x(283)	
MOTVGNVQVJEMWAGNARFADQSGLEIAHVAQAALITAFMGAFITLIEISHPDTQMGDOGILII			
SOPITLPLVYIGYOTHFASISLIEDLIGGIFGGLIWHILVPLGAKKV			
LIIFGEALISLSGIGLAFVAVAFCAVNTLAVPEPEFGPLAKIGIFPFADTVELPMHATSRAMLANAHFFLAFFLQGHLMH			
A64834	ck: 3224 len: 182	1 probable fibrial protein-like protein yobC	
1		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C	
1:		x(44)(A)x(T)(L)xx(A)(T)(R)xx(V)xxx(P)x(122)	
MITTKKSVLAFIIVVCAITSSVMAADNATLTDSSVTFNGVLAIPACTIVAAKDSVTLTPDVSAIK			

1	D71484	ck: 1871 len: 179	1 probable ribosome releasing factor - Chl	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x(68)(A)x(T)(V)xx(I)(S)(R)xx(A)xxx(L)x(95) MTLASAEKEMGVILTFEQKEGRGRTCKAHPLAVEVVEVGTIMRLSDIASISVSDMRQL
TV	S74877	ck: 7310 len: 328	1 phosphate transport system permease prot	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x(191)(I)x(T)(I)xx(I)(S)(R)xx(I)xxx(S)x(121) MEGEFRSEFEFLIAGCDITQGNLSLTDGKMLQWCRLOTRICSAGVVLIGWTAAVFTDARPAI 1: MEQESRSEFEFLIAGCDITQGNLSLTDGKMLQWCRLOTRICSAGVVLIGWTAAVFTDARPAI PSMLVAGLVLTWITITIASISROLILSVSPSLASAMALGATRMETISCTYPSA SSGIIIGATILALGRALGETMAVTVIGNSNIIASLAPGTTIPSLAQAFAVDELHIGAKMLALILFVITLGINSLAVIMV
1	A46259	ck: 9097 len: 438	1 recombination protein recA homolog - Ara	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x(200)(L)x(T)(A)xx(M)(C)(R)xx(A)xxx(1)x(222) DSQIVSTIKNSFPLSPISPLPPEPCSSPFRPSGCSRRRLYSVVFYAAKRLSKHISEEF 1: DSQIVSTIKNSFPLSPISPLPPEPCSSPFRPSGCSRRRLYSVVFYAAKRLSKHISEEF YSKALGVYDENTLVCOPDNGEMALETTADRMCRGSAVVLICVDSVALTPAEIEGE IGMOGLOARLMSQALRMSGNASKACITLIFLNOIRYRIGVYGNPESGIALKFFASVRLERISAGKIKSSKQEDIGLR
1	J00661	ck: 6434 len: 424	1 impb protein - Salmonella typhimurium pl	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x(145)(A)x(S)(A)xx(A)(T)(R)xx(P)xxx(G)x(263) MFLADINSFVASCCEVFRPDRLREPVVLVSNNGCIASAPARKALGIRMGQPFQYORQRL 1: MFLADINSFVASCCEVFRPDRLREPVVLVSNNGCIASAPARKALGIRMGQPFQYORQRL GLDPYGVGVGRILTEKNAIGINTALQALQANTAFKRNFTLIERVRENGE SCILSEAPAKQIYCSSEFERITDDAHQAVVOYAEBAAKKLGEHOYQOYTFVRSPEAVEKPCYSNAAVEKLPLPTQ
1	F71946	ck: 9168 len: 343	1 flagellar motor switch protein - Helicob	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x(42)(I)x(S)(I)xx(I)(S)(R)xx(V)xxx(G)x(285) MARTLPRQAOBDELSERKATILILVQGEDTGEILRHLDISITIEISKQIYQJONGDKOI 1: MARTLPRQAOBDELSERKATILILVQGEDTGEILRHLDISITIEISKQIYQJONGDKOI EISPOVYKRVSTYLENKLESLISYKIEVGLVAIVEINRUGQSAKITLARISSV DNKLGAIKEMTFEDDIKLNFAIRELKYADKKDSLAKTSTODLTKPLNNSSRAACQFVEMOYIGANKINDVYAOA
1	A37763	ck: 2312 len: 147	1 virdI protein - Agrobacterium tumefaciens	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x(125)(L)x(S)(I)xx(V)(S)(R)xx(I)xxx(S)x(6) MSQSRPTSSDIADVNORECVKVEGVSTRLRSAYESESFSHQARLLGLSDSMAIRAVARRIG 1: MSQSRPTSSDIADVNORECVKVEGVSTRLRSAYESESFSHQARLLGLSDSMAIRAVARRIG
1	A25063	ck: 4145 len: 147	1 hypothetical virdI protein - Agrobacteri	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x(125)(L)x(S)(I)xx(V)(S)(R)xx(I)xxx(S)x(6) MSKTRVTSSETAINQHRSLNVEGFVAVSARLSAYEVEFSYQARLLGLSDSMAIRAVARRIG 1: MSKTRVTSSETAINQHRSLNVEGFVAVSARLSAYEVEFSYQARLLGLSDSMAIRAVARRIG
1	S06883	ck: 2761 len: 147	1 virdI protein - Agrobacterium rhizogenes	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x(125)(L)x(S)(I)xx(V)(S)(R)xx(I)xxx(S)x(6) MSQSRPTSSDIADVNORECVKVEGVSTRLRSAYESESFSHQARLLGLSDSMAIRAVARRIG 1: MSQSRPTSSDIADVNORECVKVEGVSTRLRSAYESESFSHQARLLGLSDSMAIRAVARRIG
1	S36564	ck: 5828 len: 368	1 E2 protein - human papillomavirus type 4	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x(63)(V)x(P)(P)xx(I)(S)(R)xx(A)xxx(1)x(289) KMQCPKESLSEERLSAIOBKLIDHYENDSDINDISYQQLIRENALITLFAREHGITKLNHO 1: KMQCPKESLSEERLSAIOBKLIDHYENDSDINDISYQQLIRENALITLFAREHGITKLNHO FSSECEKGNSTWEPYGVGNVINDCNSMCKSTSDIVTSQIAIVTIOLRASISTPXT

[illegible]









1	JC5603	ck: 1627	len: 270	1	ESI protein - zebra fish	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(56)(A)x(T)(M)x(L)(S)(R)xx(A)xxx(I)x(198) 1: MLASRLALAOAAMLVORACALMHGDMGNMGNINIAVFSGCGMDGTDIHEAATMYHLNR LACRYLPSELEVTEDESSRGMCHWPTNINVOAKSMGARNHNTREPEAYVDEK NKVISTPTMETEDYHYHIFDIGNMVKHVMRTAK
1	T09924	ck: 1428	len: 293	1	cytidine deaminase (EC 3.5.4.5) CDA6 - Arab	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(80)(L)x(S)(M)xx(L)(T)(H)xx(V)xxx(I)x(197) 1: MKPYTPSEAEERGVGRPSDLPRLIDKAMSLARAPSTFRVGAAGVLSSGEVFLGVNEFPLPH ECSHTRCBAIAAANSVAPSKPCSGVALLICGCVKMYTESVAYNPSLCPVRA ALVDFAARGGKEFTEITVLEKMDVVSQDARTRTLDKIAKCDKVTLCYTKTKN
1	S55551	ck: 2728	len: 257	1	cytokine-inducible protein C1S - mouse	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(179)(P)x(P)(A)xx(M)(S)(K)xx(A)xxx(S)x(62) 1: MVLCOGSCPLAVEOIGRRPLMAOSLELPGAPMOPLPGAFPEVTEETPVQAEENPRVLDPEGI PAPTPALPKSKODAPDSVLPPIPVATAVHLKLVOPFRSSARSLOHLCRYNL VADVDCLPLPRMADYLRQYFQOL
1	JC5626	ck: 1738	len: 198	1	STAT induced STAT inhibitor 2 - human	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(115)(A)x(S)(L)x(L)(C)(R)xx(I)xxx(T)x(123) 1: MTLCLPESNGGGGTRTSONGTAGSDEEPSQARLAKLRELQGTGWTGSMVNEAKELKEAH IMGLPLPRLKDYLEKRFQV
1	JC5760	ck: 1381	len: 198	1	cytokine-inducible SH2 protein 2 - human	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(115)(A)x(S)(L)x(L)(C)(R)xx(I)xxx(T)x(123) 1: MTLCLPESNGGGGTRTSONGTAGSDEEPSQARLAKLRELQGTGWTGSMVNEAKELKEAH IMGLPLPRLKDYLEKRFQV
1	A41900	ck: 4304	len: 299	1	cyn operon regulatory protein cynR - Escher	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(1192)(L)x(S)(A)xx(A)(T)(R)xx(I)xxx(C)x(91) 1: MLSHINVFALVAHEGSEFTRASALHVSQALSOQITROLEESIGVPLFRSGRTIRLTGAGEVWRQ HEQVALSRHDEKLVLSAEFATREQIDHYCEKAGHPQVTEANSISAVLELIR TSISTILLPAIAIOTHDGLKATISLAPLERTAVLLRRKNSQOTAAKAFILHALDKCAVVGNESSR
1	F70721	ck: 7074	len: 164	1	hypothetical protein RV1829 - Mycobacterium	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(49)(V)x(P)(P)xx(L)(T)(H)xx(I)xxx(I)x(99) 1: MGEVAVGIRVEOPONOPVLLLEKANEODRILPIWIGOSEAAIALLEOQGVPRPRLTHLINDLIR
1	S56618	ck: 3873	len: 173	1	yjjx protein - Escherichia coli	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(1144)(V)x(T)(A)xx(L)(T)(R)xx(V)xxx(A)x(13) 1: MLIMHVVCAITNPAKIOALIOAFHEIFEGSCHIASVAVESGPEQPGSEETRAGANRRVANAAR
1	S45255	ck: 4167	len: 54	1	yjjx protein homolog - Enterobacter aerogen	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(24)(V)x(T)(A)xx(L)(T)(R)xx(V)xxx(A)x(14) 1: EALGCVNQHGTGIDIGKREGAIGVFTGKRLRSSVYHQAVALLSLFFHNAIYR
1	A70738	ck: 8373	len: 158	1	probable rimI protein - Mycobacterium tu	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(6)(P)x(T)(I)xx(L)(T)(R)xx(A)xxx(I)x(136) 1: MTADTEPVTIGALTRADAOACAELEAOLEFVGDDPMPPEAFNRSLAPNHHVYGARSGTLVGY
1	F72316	ck: 5138	len: 309	1	hypothetical protein TM0917 - Thermotoga	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(79)(A)x(T)(V)xx(M)(T)(R)xx(I)xxx(T)x(214) 1: MFVLLDSEFLGEMGANDANAEFPGVSGGLIPYRKATIVASIFVLISVYGARGCQNISS TSLIAVLSQAMVWIFSLIGIHVSSQALVAGVAGYKGNLGNKYLILKILSGWFLTPAVSGTSLFLLTSLIK
1	B70451	ck: 4885	len: 311	1	prephenate dehydrogenase - Aquifex aeoli	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(9)(P)x(P)(P)xx(F)(C)(K)xx(L)xxx(L)x(126) 1: MAIISSENFSPGPGFCKNIILKLSLQNVLLVGVGGSGFAKSLRSGFGKITYGDI TDKRLKLVKRVNEDVGVVYSPHLPVAVAFALVDLTHMSRP EVDLFKYPGGGFFDRIAKSDIMWRDIFLENKENVMAIEGFESLNHLKELVREAEELVEYLKIKRMEID
1	G39845	ck: 2658	len: 256	1	dihydroorotate dehydrogenase (electron t	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(118)(V)x(P)(L)xx(L)(S)(K)xx(I)xxx(V)x(122) 1: MKRAYLVNCSNOOLADIVFOMVYKGEVFOFTPGQFHLKLVSEAVTPLLRRPISADVNEFK VIRKRLPEFLLSCGTPMLKRLKOEYANREYVLSMEBRCMGICGACFACVCHTN ESERTSYKVCOLDGPVFAKEVAL
1	J01382	ck: 6785	len: 303	1	hypothetical 34k protein - pea enation m	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(1170)(V)x(T)(L)xx(L)(T)(H)xx(F)xxx(L)x(117) 1: MHQIEQOLPELDIVYHRCASSTFLASLDGLGSLARSLGSLALITSYLLVLSIALCMAIPGS THODPRYLALRRNLDLSEEPFRVARGVLESALCSCVSTKSTSKQRPPIGVS LHVLGLAERDCILFDIDSNTSPTIRHVLIEDORNRADOSLFSIDELVHDLIALGYSDDEDLDLNF
1	S56961	ck: 8234	len: 196	1	probable membrane protein YJL178c - yeas	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(1167)(L)x(S)(L)xx(F)(C)(K)xx(V)xxx(C)x(100) 1: MLCGLTVLILPGKDAITTOIIDFKDNIGFVNEETESALTLTLKGATMGANSFPAKLEFOCNDN
1	VVSRIIGRSTAOGRGYSAY	ck: 9413	len: 134	1	V0 protein - Miscanthus streak virus	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(118)(A)x(P)(P)xx(L)(S)(R)xx(A)xxx(C)x(100) 1: MGCYFIMGRGSEINNTIAPPPMPISPLASVCACGPRSLPLADRDQAAKMRHLSVFA
1	J00150	ck: 6657	len: 122	1	hypothetical 13k protein - Pseudomonas a	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(98)(P)x(S)(A)xx(A)(S)(R)xx(A)xxx(I)x(8) 1: MALASPCSRCCCAASABMPATACCPYGAISRTCARISIPMIAMRWITCSTATVACARSRR
1	B71217	ck: 8463	len: 108	1	hypothetical protein PH2002 - Pyrococcus	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(25)(L)x(T)(L)xx(L)(T)(R)xx(V)xxx(L)x(67) 1: MKRHLTELKSLIRPLKIDIFVYHCHLDTLCVTRLGPGFSLHLHFHDLPLVKGHSRPSLV
1	T05820	ck: 4391	len: 217	1	hypothetical protein TSK18.150 - Arabido	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(25)(L)x(T)(L)xx(L)(T)(R)xx(V)xxx(L)x(67) 1: MKRHLTELKSLIRPLKIDIFVYHCHLDTLCVTRLGPGFSLHLHFHDLPLVKGHSRPSLV

[illegible]

D71926	ck: 2840	len: 142	1	cag island protein - Helicobacter pylori (s	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(42)(L)x(S)(V)xx(L)(S)(R)xx(L)xxx(L)x(84) MKTNFKIKLFRMCLIIQMFNAPLNADNDIKDISPEDMALNSVGLVSRDQIKTEIPKTELEOK
F71925	ck: 6963	len: 114	1	cag island protein - Helicobacter pylori (s	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(80)(L)x(S)(L)xx(L)(S)(K)xx(L)xxx(A)x(18) MKRPIKSLKONFLQFHKSFNKHLDKXSLYYRLENISSIVIGFLIALFSYGAGVILVPILEFLFA
H64586	ck: 7347	len: 114	1	cag pathogenicity island protein cag15 - He	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(80)(L)x(S)(L)xx(L)(S)(K)xx(L)xxx(A)x(18) MKRPIKSLKONFLQFHKSFNKHLDKXSLYYRLENISSIVIGFLIALFSYGAGVILVPILEFLFA
F64587	ck: 2551	len: 142	1	cag pathogenicity island protein cag21 - He	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(42)(L)x(S)(V)xx(L)(S)(R)xx(L)xxx(L)x(84) MKTNFKIKLFRMCLIIQMFNAPLNADNDIKDISPEDMALNSVGLVSRDQIKTEIPKTELEOK
A64649	ck: 6457	len: 131	1	hypothetical protein Hp1033 - Helicobacter	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(22)(L)x(S)(V)xx(L)(S)(K)xx(L)xxx(V)x(93) MYSKILATSEFLNSMWLKRSLSSVPLSKRQISKVSQALIKIDANNNAIKYPPSDVNTVHNMLG
D64710	ck: 2117	len: 115	1	hypothetical protein Hp1524 - Helicobacter	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(13)(V)x(S)(L)xx(A)(C)(K)xx(P)xxx(S)x(86) MKSRIHFHIAISFVLSEFSACKDEPKSSOSHONNTKIRKNPINOANNNDIKIEHEEDEKAT
E71938	ck: 3806	len: 91	1	hypothetical protein jhp0391 - Helicobacter	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(22)(L)x(S)(V)xx(L)(S)(K)xx(L)xxx(V)x(53) MYKILATSEFLNSMWLKRSLSSVPLSKRQISKVSQALIKIDANNNAIKYPPSDVNTVHNRM
A71809	ck: 3861	len: 115	1	hypothetical protein jhp1413 - Helicobacter	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(13)(V)x(S)(V)xx(A)(C)(K)xx(P)xxx(S)x(86) MKSRIHFHIAISFVLSEFSACKDEPKSSOSHONNTKIRKNPINOANNNDIKIEHEEDEKAT
C71912	ck: 428	len: 307	1	probable outer membrane protein - Helicobac	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(187)(A)x(T)(L)xx(A)(S)(R)xx(L)xxx(S)x(104) MKKVLITSLSLSEFWLAEENRGEFLGALFPAEGSIOGSGIGKASAEALNAINNAKNSLFPN
PH0856	ck: 7474	len: 277	1	mauf protein - Paracoccus denitrificans	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(123)(P)x(S)(P)xx(A)(T)(P)x(V)xxx(L)x(238) MYSIEDHLNLSAGASVSDCKELFPOSPPAKIRIAYVLLAALAGAGVALASAGPOPLMAVUGAA
139703	ck: 4597	len: 102	1	tram protein - Agrobacterium tumefaciens	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(113)(L)x(P)(L)xx(L)(T)(R)xx(P)xxx(L)x(73) MELEDNAVTKVELPRLIGLNGLPFLDLEITLIDALHRLRYEKADELFOALPEYIKTG
JC2568	ck: 7036	len: 118	1	mray protein - Rhizobium meliloti (fragm	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(48)(L)x(P)(L)xx(A)(T)(R)xx(L)xxx(L)x(54) LOIHFPVGTGLAVILGAVIGLGLFLEFMENAPPAIFMGDTGSLGASLAPLAVAKHEIV
S27344	ck: 1548	len: 370	1	hupK protein - Rhizobium leguminosarum	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(125)(A)x(S)(L)xx(L)(S)(H)xx(A)xxx(S)x(229) MTEFLGATIDGVTARLACGSAVYANVARNPRLTFNMFGRPEEAPVLAVGVSFCGAGSV
S28677	ck: 6944	len: 243	1	hypothetical protein 4 - Rhizobium sp. (	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(93)(L)x(T)(V)xx(L)(T)(H)xx(L)xxx(S)x(14) MSLOVATVROSVAAVAPLFAKEDQSDSLSEVFTFPAAVFEGQARHIEVEVEGTLR
S26139	ck: 4831	len: 187	1	signalling protein ampd - Citrobacter fr	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(164)(P)x(P)(A)xx(W)(S)(R)xx(A)xxx(T)x(17) MLENGLVYDARHVPSPHDCRDEDEPFLVYVHNISLPPGEGGWIDALFTGTIDPDANHPF
A48901	ck: 4831	len: 187	1	signalling protein ampd - Enterobacter c	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(164)(P)x(P)(A)xx(W)(S)(R)xx(A)xxx(T)x(17) MLENGLVYDARHVPSPHDCRDEDEPFLVYVHNISLPPGEGGWIDALFTGTIDPDANHPF
S40867	ck: 3905	len: 248	1	ferredoxin - NADP- reductase (EC 1.18.1.2	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(129)(P)x(T)(A)xx(F)(T)(R)xx(L)xxx(G)x(203) MADWYGTGVKVVQWNTALSLVHAHAFVFTGQTFGLGLEDERVORAYSYNSPDPDL
A64750	ck: 8141	len: 79	1	hypothetical protein b0249 - Escherichia	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(5)(L)x(P)(P)xx(F)(T)(R)xx(A)xxx(T)x(58) MOSVLLPFGPFTBRQAOAVTTTYSNITLEDQDSHFRLVVRDEGRWVRAWFEPDAGEE
C65059	ck: 1867	len: 425	1	hypothetical protein b2775 - Escherichia	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(152)(V)x(T)(W)xx(P)(S)xx(P)(S)x(257) MOMNSYRMTLALISFSGVSDALATRIYIOIPAKMFGSNTFGLIMSTFGAIIILYA
FTTALISYSTNYITETGSLVAASWKGIVIKIRFALCGPLGGITTYTSVKSPTVNIOLISVGLLTLTALLVTSNPOSVA					

B65068	ck: 6883	len: 145	1	hypothetical protein b2849 - Escherichia coli
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1:	x{99}(A)x(S)(I)xx(L)(S)(K)xx(L)xxx(L)x{30}			
	MYTMDIESQIHEVMYHMDIVNSDKRRPRIPKFKELMENVLTQTSWTLNSRYVNSVKNV			
F64963	ck: 7029	len: 359	1	nicotinate-nucleotide-dimethylbenzimidazole
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1:	x{137}(L)(X)(S)(A)xx(L)(S)(R)xx(A)xxx(L)x{206}			
	MOILALMTALITDSTASRAQRHIDGLKPVGSLTEVLAIQIAGMPGLNGIPVGRKAVLV			
	NANTPAAIYSTIGTDPPEVYMGALPTDKLNTDIVRAITLNOPODGV			
	DVLAVGDFDLVINGVNLGAASGLPTLLDGLFSLYAAALACMPAIRIPYLPHLSAEKARIALSHLGLIEPLNNEMHLSGSG			
T00211	ck: 5937	len: 291	1	Type II secretion pathway related protein e
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1:	x{51}(A)x(S)(V)xx(L)(S)(R)xx(A)xxx(A)x{224}			
	MLEFISFGRDGLFKIDVLKRLTPNRLCYLLAGIOLVSLHFMWLOASVPELSKVSAPET			
	EKKNAVNNNTLVGVITQDNFRVSEVDIFSFAVYVNNLSGTRILPKHSHSL			
	YNAAGLHNDLAVSVNGSELRTROAQIMKQIPKLEIKITVERDGLYDAFLAVNGEN			
H64726	ck: 98	len: 216	1	yabp protein - Escherichia coli
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1:	x{8}(P)x(T)(L)xx(M)(S)(K)xx(L)xxx(V)x{192}			
	MRVSVGMVTLLNMRKNDIYKMGSDKMDVMNIFORLWELRLHFWMSDKOTEAVALTFEYVNC			
	SNDIYERPGMANNIGVLPRTVLPRTVLTPLTWTVP			
A64862	ck: 1943	len: 78	1	ycgZ protein - Escherichia coli
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1:	x{7}(L)(X)(S)(A)xx(L)(T)(R)xx(A)xxx(L)x{55}			
	MHONSVTLDGALTRFRANLHTQETIGETIETILNDGRNLSKSKSCATLRLSHAGGE			
A64882	ck: 6021	len: 262	1	ycjI protein - Escherichia coli
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1:	x{16}(P)x(T)(L)xx(V)(T)(R)xx(A)xxx(A)x{230}			
	MIKRICTTLPVLPETIMVTRPRARGAFPEOTEHIGRSLGAPLWFPAPASRESGLILA			
	DPLACIEDRHSSELGEMIAQAFELPLVTSVGETPGSGWCADLNLHCTAEPF			
	ISSDEASEKYLFLAMANLIRMBKDAIRPS			
S20452	ck: 9681	len: 271	1	hypothetical protein X - Klebsiella pneumon
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1:	x{138}(A)x(S)(P)xx(A)(T)(H)xx(A)xxx(C)x{117}			
	SVSPDPIIDMOODALIKOALIASAGRLRLCSDIEREDKVLAVAVHIEGAGFGDEGRDLC			
	NFGNAFLRDGRSDPTPLTTIYRHIDYIINNGDHALIKVSGDFGILLPDELGDV			
	AGLPRINTLRASGYDQVLDKLIMRWMLRVKKNWOO			
S01838	ck: 7149	len: 220	1	nifY protein - Klebsiella pneumoniae
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1:	x{60}(V)x(S)(P)xx(W)(S)(R)xx(A)xxx(G)x{144}			
	MSDDITLFWALAFQSLPDQDQVYDMLDQSGEITLIPERLALLTQPOLAASPSATAVMSPAH			
	MKAQPTTIOAOCEALINTLLAGRLPWLAKRLNRDPLERVF			
S22619	ck: 7734	len: 336	1	hypothetical protein - Salmonella choleraes
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1:	x{115}(I)x(T)(V)xx(F)(S)(K)xx(L)xxx(S)x{205}			
	MNNKVLMDISWSSKSGIGRFTDEISKEELRYKRCASPLAPGLAVNIFLRKKTVDVFLP			
	VISAFKADIDPSIKIVFTGNPCNDLEKILIOHGLSERKFEFVSEKDDLSLYKG			

SGLVPSLYEGEGLPVEGMAGIPVLTSLTSSLEPVAGDAALVDPDLSADATKGISRLINDELKHLIOKGLRAKRFNMO				
A53302	ck: 5663	len: 289	1	hemf 5'-region hypothetical protein - Sa
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)			
1:	x{122}(A)x(T)(L)xx(M)(S)(R)xx(A)xxx(L)x{251}			
	MSTRKLTLSRROVKTGLALTLISGSHAVAAEETLTKISNGSKRKTGSKRLVMDP			
	ATDRDHLLOQVLPDVOGDTIKNSLTLSGSHIIRKIPKIRKHSRTTEQAAEVNKS			
	PSIPSVLVSFTINPEERLGTAFKRIATIANGLIISYFHWFDNOKAHTRK			
S23906	ck: 4054	len: 135	1	hypothetical protein 1 - Shigella flexne
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)			
1:	x{129}(P)x(T)(A)xx(F)(T)(K)xx(L)xxx(G)x{90}			
	MADWVGKTVKQWMTDALPSTLVHAPVLPFTAGFTKLGLEIDGERVORAVSYVSPDNDL			
T14663	ck: 2204	len: 99	1	histone H5 like protein - Yersinia pestis
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)			
1:	x{47}(A)x(P)(V)xx(P)(T)(R)xx(L)xxx(A)x{36}			
	MKATIVGANSFSAIYSSKNDLIRDDIVETIADIVPMKRGTAEVBAKRVQSTFEELS			
T14952	ck: 2530	len: 144	1	hypothetical protein - Yersinia pestis p
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)			
1:	x{92}(A)x(P)(V)xx(P)(T)(R)xx(L)xxx(A)x{36}			
	MKPAKIRLEPQFLGTVGIGIOFVDGISVLAERFDIDORICASMRATTVGKAVSSAAS			
S70883	ck: 2492	len: 227	1	hypothetical protein 3 - Vibrio cholerae
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)			
1:	x{58}(V)x(T)(I)xx(P)(T)(R)xx(L)xxx(I)x{153}			
	MARLFEMNLSRVKRRISLTVDAFEIIFESYSAVWVNGVEIHSDSIPYLATVAVTIL			
	KTLENTVMQGVHDVSRAYLVKDYVTOILAVPSASRAIUSDCAKT			
S54442	ck: 6509	len: 150	1	membrane associated protein 17.9K - Vibrio
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)			
1:	x{35}(L)(X)(L)(A)xx(M)(T)(R)xx(L)xxx(A)x{99}			
	MKIGERERAFVLMVAVVVLITANFWSVRBMFDMHTALEMLRQLIDRANVQGEWVLQGR			
G64102	ck: 8973	len: 288	1	phosphatidate cytidyltransferase homol
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)			
1:	x{108}(P)x(S)(A)xx(W)(S)(K)xx(L)xxx(L)x{164}			
	MKOVNLSAIVLAVLACALFEPFETALALAGVAILGIEWTOFARLQKPLIRFTVTFIG			
	GKRKLAPRVSPGKMEGVITGLITVALAFIHFSSNLTLVGDRNTGFIILSVAT			
	VAISVGLDTESMFRRESGVDSQILPGHGVLDRIIDSLTAVAPPEFSYFEFVL			
H64157	ck: 4	len: 311	1	sufl protein homolog H10733 - Haemophilus
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)			
1:	x{222}(P)x(P)(L)xx(S)(S)(R)xx(L)xxx(P)x{273}			
	MRLSRROLTAKAALSTLSTVPAAPLAAASERKLVVPLLEVRGRPIVLTMOETVPLDGS			
	GYVDIPLIODEMFEFNQDLOLFKONOHFNGRLNVGIEAPYLDVARCMIRRL			
	NMSLARAYDLRINDQDEHLLLAQDLGLPRAKSVKSLVSPGRRAVLVNMKLTLYLSLAEVNAVCTRKKNKYVLLR			
H71692	ck: 1607	len: 405	1	hypothetical protein R9358 - Rickettsia
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)			
1:	x{92}(L)(X)(L)xx(L)(T)(K)xx(V)xxx(L)x{297}			
	MYTLISLFLPLSGVLTGSLNIAVAFAFLLSLITIGISFLOKQEFNKKFIKIYPRFGD			
	ILFSGNHINSFILTIVLYLTSISDSIASLFGSISGIIIFILARKLITIFLVLIT			
	ISLITGSLFVPVIAQIDPQNLSEKATYSAHRLFIWFAVANKIILIPILVGFASSKATILEGDNMAIDYRQEKIHLPLHP			

E72060 ck: 2665 len: 252 i ct470 hypothetical protein - Chlamydia pneu  
1  
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x{147}(L)x(T)(A)xx(L)(C)(K)xx(L)xxx(C)x{189}  
MOICATGVLTSRPGKHHHTLTLETPGLTFPAKOSQOTQOCYRELIVPISLGTKTLHNSRL  
KQATSEKEPEOITLAIHAKORSELLAIEFIALAEKIFLFDLSQEKSEERN  
SSEDPYHEILRLSVHPY

F72031 ck: 6017 len: 493 i ct578 hypothetical protein - Chlamydia pneu  
1  
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x{189}(I)x(T)(L)xx(A)(T)(K)xx(L)xxx(A)x{288}  
MSISSSGPDNKNIMSGVLTSTPGVPQODKLSGNETKQIQOTROGKNTMESDATTAGSGDK  
EVLIEGLALAKAIQTGEATKSLSNASTAQADQTNKLGLEKQAIKIDREERY  
QEMKAAEQSKDLEGTMDVTNVTAVSAITVIAVLTGAGLAGAAGAAAGAAATVATQITVQAVQAVKQ

S61492 ck: 5418 len: 270 i hypothetical protein 2 - Chlamydia psittaci  
1  
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x{116}(L)x(T)(P)xx(F)(S)(H)xx(L)xxx(L)x{138}  
MSRSSNKHSSRTEENSTSTMEPIADYKRIYOGEGHYHKEVILPKLLPLMNLQSEDSLVDIG  
KLLSRKIDYLTAKITVSIYAHGKROKSESSIFHFLSWTQALSKYGVINEME  
EMISPKSIGAHAKAENLCREFFPLMISCKTHKN

JG5205 ck: 8587 len: 160 i sulfur-rich protein - Chlamydia psittaci  
1  
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x{133}(V)x(S)(V)xx(W)(C)(K)xx(L)xxx(L)x{111}  
MLTGENSESEVIDLIRKGDVKNKEIVQVTLVNSVGMCAHIVPIRTSKIVQSRQIIMVY

A71529 ck: 4189 len: 303 i hypothetical protein C7324 - Chlamydia trach  
1  
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x{81}(A)xx(S)(L)xx(I)(C)(H)xx(V)xxx(L)x{206}  
MYKAAHPHSSISALPSSRGGAHDSLDLSPISLPALSVLTPEPAOSKSRVQAVANTIG  
LGHOTSSLAAYGTRFSGPGYIADQELFDEORIRQALMOCKLIRNOMLQRRD  
TLQKRAHRRQOSQITIRPLTNRLSLPEEMEHILCKKPOVITAMLDNKIAQLNQSIVLYDRPRTVIEQ

C71569 ck: 1406 len: 300 i probable ribonuclease HII - Chlamydia trach  
1  
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x{167}(L)x(S)(L)xx(W)(T)(H)xx(I)xxx(L)x{117}  
MPSSEFVSLSSLSILREOLEKGFITISIPHTVFOGRSPVSCIVYQSGKIVYQSGKTOEFVEF  
TIIDNLAHPAGAVPAISDOFASSEFVLLQAVKKSDEILIRHRAEDOVVAAV  
SIAREAFSLSIHALSEQVQIRLLKGCAGSKVORAKELIHNKGOVLEKVCITHEFTFEVLGSGNQ

B53203 ck: 1180 len: 43 i hypothetical protein 2 - Desulfovibrio vulg  
1  
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x{123}(A)xx(P)(A)xx(A)(S)(R)xx(P)xxx(P)x{4}  
MTVGLANPQANPPDHANITANLADPAGASNSLSPDMGRRA

S17812 ck: 2127 len: 192 i hypothetical protein 192 - Rhodobacter caps  
1  
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x{142}(V)x(T)(L)xx(L)(T)(K)xx(M)xxx(I)x{34}  
MDLDFEFAHVEVEEMATHYIPEARIGAAMDSDRISGFAVTTAISRLQELHAIQTLVTADS  
VTNDLSLVISEFSLV

S22631 ck: 1875 len: 166 i petp protein - Rhodobacter capsulatus  
1  
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x{68}(L)x(S)(V)xx(V)(T)(K)xx(L)xxx(L)x{82}

1: MADTGAPGGETLLFTDEQLRKGIEMAFYAGFTADPRILIDQHYGRAHRAIHFINREPG  
B58863 ck: 8138 len: 213 i transcription negative regulator ChR -  
1  
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
x{122}(A)xx(S)(L)xx(A)(T)(H)xx(L)xxx(C)x{175}  
MTIRHVSQALLTAIYAGTSLSEAFSLVATHSLEDCECARAGALDANVGSIMETAPVASE  
TPVARGLDCTCLATADAPLRNNSFLPVLVPPFRI

S36981 ck: 4069 len: 201 i hypothetical protein 7 (atpc 3' region)  
1  
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
x{117}(P)xx(V)(T)(R)xx(A)xxx(P)x{168}  
MYLAPRGVEYIOLIAPCGSRPLVPAPAIIEIGALOVPAORHORDOPVMPNSVLANLMARV  
AANVTPTPLKINGQMDVDYVQRI

S74634 ck: 2111 len: 393 i heme binding protein precursor ycf10 - S  
1  
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
x{180}(P)xx(T)(V)xx(L)(T)(K)xx(V)xxx(L)x{197}  
MLPDKRLGSPNHRFHMLVROALDANOVSAQALIESIEKTYFDGKRIAPDSQGVNTYNF  
IFIPLTVOILTKNLVAFVYOHFRVDIVAMERKHQOEETIEHYEEFARXETLEI  
KQLSENQPLNOKERIHQELKRAEBELIRQATNSQGIIVNLADIAGVAFVVLIVFRKSIITQOYLSQSFALNDIRKVFIF

S76728 ck: 9759 len: 187 i hypothetical protein - Synechocystis sp.  
1  
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
x{157}(P)xx(S)(L)xx(V)(T)(R)xx(P)xxx(A)x{114}  
MUNKSQVILSGVYLAALAGFTTPADAEFFQIKGNILSLQGLGHMIVLPDPQMSFDGLAIGR  
PAPMPTIRGLM

S76937 ck: 9556 len: 312 i hypothetical protein - Synechocystis sp.  
1  
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
x{139}(P)xx(V)(T)(K)xx(V)xxx(V)x{157}  
MMKRLILITGGGYIISVLPFTLLAGVYIVADNPMFOONSIAECCOYETENVIRGDCRKEDL  
VDLVNDEVYRAFDRAVYIEGHEKRNRYIHIRVAKYATVILPGLSFMCKRPVNV  
GLEDANLSKLEICAEIRKRYLNFVYLEAPIGEDDEKRDYIVSNORILSTGTPDWSLGRIOELIKYITLIRNSVSNV

S74351 ck: 4892 len: 333 i hypothetical protein s110071 - Synechocy  
1  
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
x{100}(A)xx(P)(I)xx(I)(S)(R)xx(M)xxx(G)x{1217}  
MKRLVTLFGLGKRFVCGGLMELKTKIKVOEICPATIVTYKQREPNTLFLDDILSGDDLGQV  
VEDLFGIFNPAKRYVLYDSAEYMAOGCGVSEGGLOPMEMAGCGAFSSVNGGSLDY  
LDPGNCKINGVYALDYDCQRLILGVVSGRMRVNDNLAEYRENNILIELRLVLLDINEFPDHYHNSFGOTIEPLQRRIAMLK

S74733 ck: 376 len: 156 i hypothetical protein s110931 - Synechocy  
1  
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
x{93}(I)xx(P)(M)xx(L)(T)(K)xx(P)xxx(G)x{143}  
MASTRKILQKFTILASGLAFISMTVMPLVLRGNANOTEGSOGTPOOPTAADLERLKEVAG

S74723 ck: 8425 len: 125 i hypothetical protein s110939 - Synechocy  
1  
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
x{80}(A)xx(T)(L)xx(L)(S)(K)xx(L)xxx(I)x{29}  
MEILDHYGLNLSVITRFVLESVSLCIIILGLVTLQMVAVAGDRRTMTNTRKFFNSIRLOFGL

S75160 ck: 4206 len: 215 i hypothetical protein s111635 - Synechocy  
1  
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
x{156}(I)xx(P)(L)xx(V)(T)(K)xx(M)xxx(V)x{43}  
MVRFTSLRKPIVIVDGRPLSREGILAIACARSSNGENPNNTKRLQPCIRBGHSIREMVDN



YESVRCDOATOKEREHREIALARKIKEMKEPFLVAALAEW

S77220 ck: 3780 len: 285 1 hypothetical protein slr1442 - Synecocyst

1  
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{100}(A)x(S)(L)xx(A)(S)(K)xx(L)xxx(G)x{169}  
 MLTVHRADROAIIIFLCISGLIARDMSLSPLAVAVLATALVOSLFTINOQNGFSWESLDT  
 YACIDLRNMAIGMPSPVPHHILNGLSVLPFLMTPDRSPINPARRSRLIMATAI  
 AIFSLIIQHYHFLPTALFMAFLPCLSPVLDRQFAPRFQWQKKNPVSN

S77428 ck: 4795 len: 79 1 hypothetical protein ssr2060 - Synecocyst

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{57}(M)x(T)(M)xx(L)(T)(H)xx(A)xxx(G)x{6}  
 MWSLGNPLPTIKAKLINDSKIAIYTAISIIIAVAVLVQLIDFREGVAPAPMETMGLI

B41838 ck: 3600 len: 384 1 Vans - Enterococcus faecium

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{84}(V)x(S)(L)xx(L)(C)(R)xx(L)xxx(A)x{284}  
 LVIKLNKNNDTSKLERKRLHRIYALIVAVLVILIRSKIRKLGDMWLSLENKYDLNHLDAK  
 SLIDBAPMPYDOKRKRVHITLDKAVRELDIEFFETIRNLQITILTKTIDLY  
 YMLVOMTEDEFYPOLSAHGKQAVIHAPEDLVISGDPDLKAVFNNILNMAAISDSIIIDITAGLSGVYSIEKNTGSIPKXKLAI

F69937 ck: 3316 len: 224 1 conserved hypothetical protein ypfG - Bacil

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{88}(I)x(T)(V)xx(I)(C)(R)xx(A)xxx(G)x{120}  
 MYNADVLAFGASDVEIEMGIGITAKYKQKRYMIDDLIEALSNGVSLRKEEAAARILGA  
 KSGFIRKSDSVSTPLTNGVIEIIEVEAREKLYGKEAGVEAYSEFPNGC

S39739 ck: 3644 len: 394 1 efflux protein homolog ynfF - Bacillus sub

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{169}(L)x(S)(A)x(I)(S)(R)xx(L)xxx(P)x{209}  
 MKOLKNSYLLYAGALISMGGCYLPPALLISTYHDDYVTSQVIVRSIPWFOPLGLVDRKI  
 RILOFVPVIOSENIRKAFQSPREKREKTSFYNAFTMALMGVYSEFP  
 TVSRFLDGEIGIFITIFIGRIGIGALVASKMGFNNRLFTVLSIVSLALELTFPIFAVSIAALIEFIAMEXEVLAKRVQ

C42365 ck: 1150 len: 208 1 flagellar assembly protein fliH - Bacillus

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{170}(A)x(T)(I)xx(M)(S)(R)xx(V)xxx(L)x{122}  
 MARKEEDNRISQANSHENIRROEQEKNMAEKKILIEAKAGGEGVALGKAEMKOYAE  
 TPRGRVDSVDTQLMKDKLTLTALFAGAAE

C69820 ck: 5172 len: 83 1 hypothetical protein yhaY - Bacillus sub

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{37}(L)x(S)(L)xx(F)(T)(R)xx(V)(L)xxx(P)x{30}  
 MTLISLEOPFLKAKKAWLSRLIYWLVDRIKSPFSLSLRPFYKNHSGTIPCLTNALTLT

E69976 ck: 1724 len: 36 1 hypothetical protein yrkG - Bacillus sub

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{4}(A)x(T)(P)xx(V)(T)(R)xx(L)xxx(L)x{16}  
 MIVKAMTPREVTKKSLTNRNLVLFMVCWKMLTVIGR

S39656 ck: 2631 len: 252 1 lipoteichoic acid biosynthesis protein dltE

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{147}(V)x(P)(V)xx(A)(T)(K)xx(L)xxx(S)x{89}  
 MKMTNNTVLTIGSGAGIGLILAKLLELGNVETICGRSARLAEAKOOLPNIHTKQCDVADRSORE  
 IEMAPPMVDIOLNKSNDKGLTIRGISSEYVYFLDGLKEGKEIITNERVEGLR

DATRADYDLRFEOONTQEN

S72921 ck: 2352 len: 86 1 hypothetical protein B2168\_C1\_172 - Myco

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
 x{37}(V)xx(T)(L)xx(V)(S)(K)xx(V)xxx(V)x{33}  
 MSTNGPSARSACKSVRDIGSEGOQPTOLFVAEVALIMRVSKMTVRLVHNGELPAV

D70669 ck: 7905 len: 381 1 coenzyme F420-dependent N5,N10-methyl

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
 x{109}(R)x(T)(L)xx(L)(T)(R)xx(A)xxx(I)x{256}  
 MGRLEFNDALVHSLPLPTLARSMAATYAGDSYVGDHINALVPSIATSEYLGIAR  
 LEDPEPRGKPEIETWAAGPRALNRATGIAATPIYVPRSPYSRALLEVRSAA  
 SDAGRDPSITPAAVRGLITGRNNDVEELESYVKMTALGCVREANARIGVHPAGDFSCVODIIPQIMDKQTVLSYAKVP

C70828 ck: 2874 len: 97 1 hypothetical protein Rv0463 - Mycobacter

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
 x{36}(L)x(T)(V)xx(W)(S)(R)xx(L)xxx(V)x{45}  
 MRRASTDTPOLIMGALIGVATGILMLAALISGDLITVSQMSRVVLLSLVAVCGAAG

D70510 ck: 2571 len: 300 1 hypothetical protein Rv0519c - Mycobacte

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
 x{15}(M)x(P)(M)xx(L)(T)(R)xx(L)xxx(G)x{269}  
 MLRRGACAGTDRGIMTDMADLTRLRLRAGAGAGAGVAFGLVDPLEPOAARPEFP  
 SHVGLFGMSMGYGLLIGARLGPATGICATSPALFTSGTSPGAFDSYDLY  
 QHSYVGLPALNSIPLRVDCGTSDFRYATQFVNOIHOHPAGSFGHGDASYWREQLPGLAMMAS

F70547 ck: 3126 len: 294 1 hypothetical protein Rv0547c - Mycobacte

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
 x{187}(L)x(S)(V)xx(A)(S)(K)xx(L)xxx(S)x{91}  
 MKRPLRMVTEBITLACGRPRPISPOLINPAPVDLIGKRIILIGASSGIGAAATKQFGLH  
 TWGVSEASAPLESYVNAKSKALNSVRSRIETEMSGOSGHSTIYELVATPIAAT  
 KATDGLPALTAEEEMVYIARPRPALNARVAVANALDISGPRVNAIQRNRNEOLNP

C70708 ck: 4949 len: 259 1 hypothetical protein Rv0776c - Mycobacte

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
 x{111}(M)x(P)(L)xx(A)(T)(R)xx(L)xxx(P)x{132}  
 MYRVEDLMAKRNPTGYAAYDADGCTGVAGAAARDASVLAALRPVVGDCLVAFDAPLVAN  
 AGOPDWVSLRQVYTAQKRSIDIRAEEDPIDAVCAIVLAQRPRADVTIYGDFTT  
 GYIVTSPLPTDFERTAPDAGRARRAR

G70716 ck: 883 len: 282 1 hypothetical protein Rv0953c - Mycobacte

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
 x{100}(I)x(T)(L)xx(L)(S)(R)xx(V)xxx(V)x{166}  
 MHGVLVFTSDRDTIAAARALESHTYVDEHHPKROAHPTTGDSLSDDDDYMYMT  
 IPIVLVGAAGTEKKNFARIASAGVITTPRDVYDIDEPVLLDIDIAAGRGRLPOIV  
 ALDVKLPVDPDLARMAELGTEVLFGMPDRSADDAAYVERLAALACCV

D70897 ck: 9641 len: 103 1 hypothetical protein Rv1102c - Mycobacte

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
 x{11}(A)x(P)(V)xx(L)(T)(R)xx(L)xxx(L)x{76}  
 MRPHIAQLDRARVLLITREVVRPHLNTVAPITTVIRGLATEVPAVDANGLNQPSVVS

F70958 ck: 4095 len: 212 1 hypothetical protein Rv1377c - Mycobacte

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
 x{133}(A)x(S)(V)xx(A)(T)(R)xx(A)xxx(L)x{65}

```

1: MGIDPEALYRGSPGSGLEPIITTPMPDITAPADNITGNHTGWHGVLDIGCGLGDNATILARN
GMIDLESLEPATVRELDGTEVEKAFNNVNRQRGGS
C70763 ck: 497 len: 148 i hypothetical protein Rv1558 - Mycobacterium
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{17}(A)x(S)(P)xx(W)(S)(R)xx(A)xxx(M)x{125}
MPLSGEAPSPLDWMSREAOADTYMKSGGTEGTOLOGKPVILLITVGAKTKLRLPLMRVEHDOYZ
C70541 ck: 1907 len: 170 i hypothetical protein Rv1577c - Mycobacterium
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{149}(A)x(S)(L)xx(V)(S)(R)xx(A)xxx(L)x{5}
MAELRSGEGRTVHGTIVPYNEATTVRDGEFQEFQEMFAPGAFRRSIABRGHKLLVSHDARTRPV
1:
A70639 ck: 9160 len: 181 i hypothetical protein Rv1957 - Mycobacterium
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{135}(P)x(T)(L)xx(L)(S)(R)xx(P)xxx(G)x{10}
MTDRDADDDLDORVGARLAKAQRIRDLRLKQAANHRAPAGQGLTYDLEFEPAVDADPATISA
RGTP
B70758 ck: 1836 len: 317 i hypothetical protein Rv1996 - Mycobacterium
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{111}(V)x(T)(L)xx(I)(S)(R)xx(L)xxx(L)x{190}
MSAOTINIGYIVGDGSPCHTAVMAQOMRNVALRVQVVPVITAPGMAFEESRFQEOAOK
GSPTSGLAIAEAFDEASRRGVLDVLAHANSMDKPLDFPRLNAPAIEMRLDEDOEK
MLARRISGMODRYPDVVHKVYCDRPPARLLELAQTADLVVVGSHRGFGPMHLGSVSAVNSGAPYIVARIPDPAVPA
A70941 ck: 7719 len: 239 i hypothetical protein Rv2018 - Mycobacterium
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{144}(L)x(T)(L)xx(A)(S)(P)xx(V)xxx(A)x{209}
MAGQDELRDPLVLTALASRLIYPRATITATWDGIERRRANPANOQGPITLALPHPIGSHR
LDNATA
1:
YCDANVYVLDPRRGITGGPVEDGSGVAVDVGFLAKAQTQAVADIGVTFQDQDN
D70943 ck: 7125 len: 324 i hypothetical protein Rv2037c - Mycobacterium
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{140}(V)x(T)(A)xx(L)(S)(R)xx(L)xxx(P)x{168}
MALVSTARVPLVCGGGGSGRIGLAVGAVDALADAGYRFPYVAGSSAGAIVASLVALQTAGEPVTRI
SSAIPFEFEPVAVRGATVWDGGLLSNFPALVDEDRTAEGEVRWPTFGIRLSARPGIP
TRVQGVSLIGIAIETLVSNOBNAYIDPCVTARTIEFVADVSRIDEDITAEOREALYQGFQAGOKFLANMYADCLADCGPPT
H70863 ck: 4861 len: 419 i hypothetical protein Rv2449c - Mycobacterium
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{183}(V)x(T)(V)xx(V)(T)(R)xx(L)xxx(A)x{309}
MTATPRREDIVATGATGFCVKTATREYATAGAGDARIALAGRSQRLVALAREALGESACWPTLAD
GGFSGCGTASMLEVLSASNDPAPAROLSDPFLSSDRABEELGPODPLSRRCR
FLABELAGVMTAGIIMPTFTRIVRSNALLDWAIRKRRIEISIMSVSTLAPVAVYVGGGVGNAMFGLASRIIRLLRGLYARVVF
1:
E70572 ck: 8454 len: 273 i hypothetical protein Rv2622 - Mycobacterium
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{185}(P)x(T)(V)xx(W)(S)(H)xx(A)xxx(L)x{12}
MANRGNAGOPLPISDBDDHMOGHMLLRLGRKRVLRPGGVELTRILLAREAVTDADVLELAPGLD
RALKVNAARPLTVAREMHLAHLGHLVVEHVTATSMALLOPSRRIYADEGLGLALRFGAG
NLIHRAARRVILMHTFRHRRLTAVAIYAHKPHVDS
F70572 ck: 9524 len: 297 i hypothetical protein Rv2623 - Mycobacterium
1:

```

```

1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{100}(V)x(T)(L)xx(M)(S)(R)xx(V)xxx(V)x{181}
MSGNSSLGTLVIGDSEAOVAANRANADBLRKLPLVLHANSPEVATMLEPPLPGVILRW
IAFDENSRNRVDLVAHANSVDVSVSMPEDIPATQSAEOVLARLGAQBERIFN
VALTRVYVRQAPARQLVONSEAOVLVVGSRGSGIAGMLVSGEIVAOIARIPVYARESLT
G70965 ck: 4149 len: 177 i hypothetical protein Rv2651c - Mycobacterium
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{156}(A)x(S)(L)xx(V)(S)(R)xx(A)xxx(L)x{5}
MSILFETALRPGEGRTVYGVYVGEVTVIRDIDGDEFREMFAGAFRRSIAERGHKVKLLV
E70885 ck: 9649 len: 308 i hypothetical protein Rv2859c - Mycobacterium
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{56}(L)x(S)(P)xx(A)(S)(R)xx(V)xxx(A)x{236}
MDLSASRSGDPLRASPRLSRSPVSDGDPRLRASPRLSRSPVSDGDPRLRASPRLSPILGA
RGTOVANAAGTIGHLEPDILGHSGHNRGNGVFTIRHTASCTRIALGICFESAD
VPCTHHQALDQVGEGLVSAVDVGYLLELPGDTFLVAVOMPEKSIDLRLFKALVDAASGYAQRQSAEPR
E70669 ck: 5693 len: 418 i hypothetical protein Rv2953 - Mycobacterium
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{83}(L)x(T)(V)xx(Y)(T)(R)xx(L)xxx(A)x{319}
MSPAEREFDLVLGAAGFGSKGLAEHLAELGSGTATIALAGRSSELRGVRMGLGNADWBLI
QRMVSGSVATYSEANRTIASDPEARLTDITLTTRGAEPELGADPDLRSPG
RDLAPELAGMTGFGVQAPFNTRIVARSNALQEMAYGGRFYSSETMSLGSMAADPILAAVGTGVAOTIGLGNKYEDRLRRLVE
B70651 ck: 3649 len: 118 i hypothetical protein Rv3073c - Mycobacterium
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{92}(L)x(T)(P)xx(A)(T)(R)xx(V)xxx(A)x{10}
MYREHRRVANAAYVEDIDPDQDQARVLVDRINPHGIRKNDQVGLWCMDVAPSKELRMYNHQ
D70645 ck: 2556 len: 332 i hypothetical protein Rv3131 - Mycobacterium
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{35}(P)x(S)(L)xx(F)(S)(R)xx(A)xxx(S)x{281}
MNTHEPDAETVRVLTAVARBSINHTQONRRVCPSTSLERPDQMLRSTDPDGRLLISC
RELTTMSGRYGVAVPARNEPSPDSAPLIRGLAGELQSPSDVLPADQGAAIL
ALQETEDDRLARLAGEAASIVLLATAMGLACCPITPELIRAKTRAVRAVEVAGAGYQDMLLRVGMADINADPLRPTRRRLS
E70951 ck: 8567 len: 355 i hypothetical protein Rv3200c - Mycobacterium
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{33}(A)x(P)(A)xx(I)(S)(R)xx(V)xxx(V)x{306}
MAGSWRLRLNKLKLAOPGLVGLRLRIPORRASPARYISRVAVAVALLTLAGIYYVRD
VDTDSGVLERAAAGLVTHGATSGATSDVLRAGTGHASITVATSRDQTAIVLT
ARIRAKAKIVASIRAEHQHLRQSGADTVVSVSETAGRLGATITTPSVVENIEDLIPERALAAVEVEGAEEVGGSPHRLR
C70561 ck: 48 len: 350 i hypothetical protein Rv3626c - Mycobacterium
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{82}(A)x(S)(L)xx(M)(T)(H)xx(A)xxx(G)x{252}
MTGASLELDLNTVDMFASVGRLLARHPSTETRRVDELTVAAEKAPPRVDTGLIA
SGYMSQALNLTLEFEPDDIGRIVASRLADIRSGTGTDSEVNSBGILGLVRAVOS
EPORKMALDQILVLTGLEHGHVNDVAVGPVAVATIRRFDRRRHNPLORLVRLALGDAKLISOYTRGKAEVDHVVDRA
H70850 ck: 8984 len: 254 i hypothetical protein Rv3912 - Mycobacterium
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{153}(P)x(T)(V)xx(I)(S)(R)xx(V)xxx(L)x{85}
MSAADKQPDHSDADADPPLTELDLADGLADDAATARISSRVASDPQAOQILBALRVVRD
PGGPIIDPSPRTSCSLGCTYPASTVLAQNPIDIDARVAVLIVPADPDLKLAFA

```

VAPRCSADTGLLASTVYVRA

A70565 ck: 5027 len: 226 1 probable cutinase precursor - Mycobacterium  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{112}(M)(S)(A)xx(S)(T)(R)xx(L)xxx(S)x{98}  
 MTRPQPSGRAPRAGARARISLVAPRAATLLTPALAPASACCPDAEYVFAAGTGPGEIGF  
 1:  
 ALTPQFSGKTINLGNNDPICSQGNRRAHLCYFGMTNQAARFVASRI  
 D70696 ck: 8388 len: 326 1 probable dtdp-glucose 4 - Mycobacterium tub  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{142}(P)(S)(P)XX(A)(S)(K)xx(L)xxx(L)x{168}  
 MEILVVGAGFQSSHLDESNHGMVYLDKSSRNANRNOGRSHDRAAFSGSVTDGOTIDR  
 FGVRQKAGRGALIPRLVROGINSGLTFLGAGSKTRDLYVSDIGVAYNLVLRTP  
 1:  
 TLNGQAINFASGKDTFVRDIVEYVADKFGARHNDARPEVDPDISLANSIQPVDELWIDIDYIMWAKDQCPYEDDPS  
 B70986 ck: 9024 len: 203 1 probable isomerase - Mycobacterium tubercul  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{163}(I)x(P)(V)xx(W)(C)(R)xx(L)xxx(T)x{24}  
 MTRSYRAPPRIERYVLLNDRGATVADKATVHTGDTPLHLAFSSYFDDLLITRRATRTM  
 1:  
 LTRKGPAPMPVADDCRLFKRAHGN  
 H70760 ck: 2319 len: 202 1 probable lipoprotein signal peptidase - Myc  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{42}(V)x(T)(L)xx(V)(T)(K)xx(A)xxx(L)x{144}  
 MPDEPTSGADPLTSTEEAGAGBNAPAPRRLMLLSAVAVVLLDIYTVAVAVOLLPPGQVST  
 1:  
 FGFDDTVGRHADGDTVGRKADG  
 B70763 ck: 4418 len: 397 1 probable membrane protein - Mycobacterium t  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{141}(P)x(T)(P)xx(I)(S)(H)xx(A)xxx(A)x{240}  
 MGSIVTGTGKGMVVRSEVFTIDIGDLOLAQVTLIDKRAIQPOLVALLPDEIASQINRE  
 1:  
 NATTFNDODGATYDILLINGIALALSLIILIMITRISVALVGVVALSLGASFG  
 LSLVWQHLLGIQIYVLAVALVLLAVSDNLLISRFKEIGAGINTGIRAMAGTGVYTAGLVFAATMSSEVSDLVLCG  
 A70614 ck: 5098 len: 287 1 probable mma2 protein - Mycobacterium tube  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{155}(L)x(T)(I)xx(L)(T)(R)xx(M)xxx(G)x{106}  
 MNVDLPHEEDVQAHYDSDDFRLFLDPOTYSCAFHEREDMTLDEAQIAKIDALGKLGIDPGM  
 1:  
 VDHGLPLTLMARFKFAVETFEPOGPTLEMEVDEOSATGPTIPROSIOPIHYA  
 RTLDLMAELQEHKSEALAIQSEEVYERIKITLGCALFRVGIIDVNOPTTLAK  
 S13338 ck: 4079 len: 256 1 hypothetical protein (insertion sequence IS  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{68}(L)x(P)(L)xx(V)(S)(K)xx(A)xxx(I)x{112}  
 MAGVITASESWAPRSGISFQSKLYVLRREGADAVRGRWWSPLIEDRALLVAAYVRNLTN  
 1:  
 GGYPGTGLVYPRRERKFAVETFEPOGPTLEMEVDEOSATGPTIPROSIOPIHYA  
 LKDDGVHNAHLGIARHNLALTG  
 PC4110 ck: 2351 len: 137 1 transcription regulator homolog - Streptomyc  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{57}(A)x(S)(A)xx(F)(S)(R)xx(A)xxx(G)x{64}  
 NHAARRLVAAPYRSGQAOVPRSVPEVGERFAATREMLARHLEGPITLLELARAASARFST  
 1:  
 PN0644 ck: 9954 len: 66 1 hypothetical protein 66 - Streptomyces coel  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C

1:  
 x{127}(P)x(P)(V)xx(A)(T)(R)xx(V)xxx(I)x{231}  
 MAKVTRDDVARIAGTSTAVSYVINNGPBRVAPATREVRVLAHKEIGTRPPRVAQANASRR  
 S44233 ck: 6765 len: 358 1 strip protein - Streptomyces glaucusceus  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
 x{136}(P)x(T)(P)xx(V)(S)(K)xx(V)xxx(L)x{206}  
 MRLTTHGQVGLGVNMAPVLTAAHQVGTGIDSGISFVRLGALADVDPVGLALDLRDVYDTL  
 1:  
 RSDIYLNIVGRVAGRYVTLSDGTATVATKATVATKIGSTIRTMKRNINGLLS  
 ENIGIEENKRTVAIEAANAAGSVLSGTDNDPSYRDFTRARTELGRATWTPIDGAQOLAREYRAGTLRAAFEHDF  
 S17674 ck: 4817 len: 291 1 serine-type D-Ala-D-Ala carboxypeptidase  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
 x{23}(A)x(P)(A)xx(V)(T)(K)xx(I)xxx(G)x{252}  
 MRLRRAATVITLQALAGTIGATPATVATKPTIAAVGAYAMNNGTGTTLTKAADTRRSTG  
 1:  
 ANSTERRHLTKIASSAMKSTFRIVVKTATVATKIGSTIRTMKRNINGLLS  
 SYSAGIGVTKGSGPEAKCYLVFAATRGKTVIGTIVLSTIPARESDARKINNYGAL  
 S17717 ck: 5969 len: 280 1 RNA methylase - Streptomyces tenebrarzu  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
 x{40}(A)x(S)(A)xx(A)(S)(R)xx(L)xxx(V)x{224}  
 MPHPAPGDPDPDPRIAEVADVADRSRRKNEKKSALIFIKVIFITFLITVGLMGCTQ  
 1:  
 VANDVRVRLMTGVGEVETVTLILKTVPCIEAOGRGQMDLIDALISPLVAVSF  
 TKSIGRSKGMNTYSANFDAMLENRPDVEQLERNELYFVKNA  
 C64251 ck: 2836 len: 385 1 hypothetical protein homolog MG464 - Myc  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
 x{128}(V)x(P)(I)xx(A)(T)(R)xx(L)xxx(V)x{241}  
 MPRLAQTNEIKITENPFWSAVVENNKNKNEKKSALIFIKVIFITFLITVGLMGCTQ  
 1:  
 KMEVQKRIAEINAKKRGALDQSKRNROLEINIKSNHNSKSAAPVQVFTLPI  
 FLIIRYITTLRPIKAILLENFMDLSKVPLTEIFENFTTGMPIFLIYVLPVQFSQKLPQWASKNENAKHAKHSQKSIQOLN  
 I64230 ck: 1048 len: 248 1 sensory rhodopsin II transducer homolog  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
 x{46}(A)x(S)(A)xx(F)(S)(K)xx(L)xxx(L)x{186}  
 MLVFEILGFTALYRSTPAALLYOARNSIDSSFNKAKFANALANSANQESKSTINMLD  
 1:  
 DVTMTTYAVSGGHALILITTYVFSFLISKKGLIFRSFISTEQDLADHVNDIIR  
 YPDLSEEVITADHD  
 S73488 ck: 5866 len: 385 1 hypothetical protein K05\_ort385 - Mycopl  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
 x{128}(V)x(P)(I)xx(A)(T)(R)xx(L)xxx(P)x{241}  
 MPNLNKNHKEKTLTENPFWSAVVENNKNKNEKKSALIFIKVIFITFLITVGLMGCTQ  
 1:  
 KMEVQKRIAEINAKKRGALDQSKRNROLEINIKSNHNSKSAAPVQVFTLPI  
 FLIIRYITTLRPIKAILLENFMDLSKVPLTEIFENFTTGMPIFLIYVLPVQFSQKLPQWASKNENAKHAKHSQKSIQOLN  
 A36149 ck: 5138 len: 237 1 spiralin - Spiroplasma melliferum (SGC3)  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
 x{40}(A)x(T)(P)xx(V)(T)(K)xx(I)xxx(G)x{181}  
 SLAVFGSAVFTSVACNCKEENKSNRSLKRTIAAPRTAASPRKAVTPEIKTALANVLA  
 1:  
 IONIAFNAGASDFEITNNGAEDYEAFAAKEVEVIVKANDSANSISGOFRKAKVAT  
 APTF  
 H69019 ck: 3145 len: 157 1 conserved hypothetical protein MTH1148 -  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
 x{3}(V)x(T)(A)xx(F)(T)(H)xx(V)xxx(I)x{138}  
 MIRVATAECFTIGFAREIHAYSMGYPGYSWSDSVVLAAGLITPLTSLGIRSLIKTEPPEP

1	D69049	ck: 1234	len: 355	1	conserved hypothetical protein MTH1371 - Me
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1		x{153}(P)x(P)(V)xx(V)(T)(R)xx(V)xxx(L)x{186}			
	1:	MNLKVAKWEIERSKRFILFQLIIVLITLTFMFGFDVLESGEALPPLTGFALGVDF			
		NLVVDSIVGERKRTGEALLAMPVRSSEIILKCSLSTIVLALQIGVMNILLIASSG			
		FHMVNPIGAFYVASSAPVGLTALISYARNREAGIGITLAIYISAAYLIAPALAYMAGSSGSVSPMTLTIKMISGOAMGLPTSF			
1	D69065	ck: 9991	len: 249	1	conserved hypothetical protein MTH1489 - Me
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
	1:	x{186}(V)x(T)(L)xx(A)(S)(R)xx(A)xxx(L)x{147}			
		MRIGFSLFMEPLLEIILQKADDFEVLCEGYPWRRLDGDGSLVFEFSDLEVLHAPT			
		YDGVISRRILLEVR			
1	G64362	ck: 6767	len: 406	1	2-isopropylmalate synthase (EC 4.1.3.12) -
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
	1:	x{160}(A)x(S)(A)xx(A)(T)(R)xx(I)xxx(I)x{230}			
		MTKVLFMFLENSKAVCPNPLDKDIDYITLTLRDEQTPGCTTKEQRLIETARKLDELG			
		VHKAEEAGADRVHIDTGCATPQSMEEIKLEENKKAHIVGCHNDGFAVI			
		NSIYGLIGAKAVSTVNGIGERAGNALEELMALTVLDVLDGLNEVLPELCRMVEYSGIKMPKRNPRIVGELVFAHESGIDHVA			
1	C64456	ck: 7004	len: 167	1	hypothetical protein homolog MJ1251 - Metha
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
	1:	x{3}(V)x(T)(A)xx(F)(T)(H)xx(I)xxx(I)x{148}			
		MITVATACEFTHANIGLTIHKAAAGYEDEFKTLFSEEDLTKMKNVRYISAMFVPSIIGVEKLDI			
1	D64423	ck: 9981	len: 329	1	hypothetical protein MJ0988 - Methanococcus
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
	1:	x{47}(A)x(S)(V)xx(L)(S)(R)xx(L)xxx(G)x{266}			
		MELLEVYKREVEFICHNDPDAVSCVATIKYLAQOLPNKFRISADSVSKSLRNLTIGERV			
		KOISFOKILYLLOSQESDVKRTHLACRKRMTREDRKRLTSLVSSHASCAT			
		IVSIGADVAVVAVAKKEKIRVAKCRKAVSVYHGLMEKIGELGSGGSHSANGLINAPYDKSKSEKVIKCVLNLCKRVE			
1	D64452	ck: 2920	len: 299	1	hypothetical protein MJ1221 - Methanococcus
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
	1:	x{108}(A)x(T)(M)xx(I)(T)(K)xx(I)xxx(I)x{175}			
		MMFEGERRKRIIIPKFLVEILVEIKNNAAVSIISIEPLKTSIEDGIIITCNMDARDAEKIVLEI			
		LKSGIIVCAFIPLSEFVSKSEIVLQYLETSEIISATVATIAIGISGMSIASGKE			
		YEIIGVITDVSILIPALLMGMLATMDLYITIFELLAINIVLVDGVIKVKYKINQIKTY			
1	F69396	ck: 9110	len: 437	1	acyl-CoA dehydrogenase, short chain-specifi
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
	1:	x{189}(V)x(S)(A)xx(A)(T)(H)xx(L)xxx(L)x{232}			
		MMWPKRLDDELSEDRMTKEVHFAFEVIRPAIEIDRRPPEBRKAPGSPYIKWQIKKIGY			
		DEWVISGOKSAWYSPVATHGGIHAOTGSGISIAEGFICVITPADACVKKGRPD			
		MLGMRRDPQSELFFDGVRIPEHIVVAFPIGFVFDQLCLTSCMGAFVAGLRACFEALHYAQRQVGPIYKHNVALKXE			
1	G69440	ck: 1599	len: 136	1	conserved hypothetical protein AF1528 - Arc
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
	1:	x{17}(L)x(T)(L)xx(M)(S)(R)xx(V)xxx(V)x{113}			
		MDVNYALLQTLTILMSKVPIMLVSLISNAPESLPQFSRVEGFINNLIIRFSKLKSGFALIASFLIHF			
1	A69470	ck: 6749	len: 61	1	hypothetical protein AF1762 - Archaeoglobus
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
	1:	x{24}(M)x(S)(M)xx(L)(T)(R)xx(A)xxx(V)x{21}			
1	D69509	ck: 4730	len: 270	1	hypothetical protein AF2077 - Archaeoglo
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W			
	1:	x{135}(V)x(P)(A)xx(M)(S)(R)xx(L)xxx(T)x{119}			
		MKRLIILVILVGLSSGNFREYGGAVANADNNSSYIGFCEPMTLYLANGDSYGI			
		LDTIAEFEEVLSIAASDPSFATHHPCGARSCTKRIWVDVEGTEFMDVTIATKHN			
		PAGKOFETSPGSGYNLNDGAIEIKGIYVNPITVAKVR			
1	T08352	ck: 7002	len: 148	1	hypothetical protein H1537 - Halobacteri
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W			
	1:	x{40}(L)x(P)(M)xx(I)(T)(H)xx(L)xxx(L)x{92}			
		MNSSPETHDFAMFPVGFDEGCRILADREXVSGFELAPMLRTTHQGLTDAEGWIG			
1	S73242	ck: 4681	len: 291	1	hypothetical protein 38 - red alga (Porp
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W			
	1:	x{116}(L)x(T)(A)xx(I)(S)(R)xx(I)xxx(S)x{159}			
		MTFASRKTELKPLKFPDPVYSIEIIOELAVORFLQWVRPATLMAGITQPLLVLF			
		VTMLSIASFLLPGHIELLALIVLVNPELFSSTALPLVFMPPMQLIASLNPIS			
		VAIEGIRIYSNTDWNFTESVIRKISWDISLGOITISLLELDVIGAVISNLIKARLN			
1	S20471	ck: 4134	len: 86	1	class V zygote-specific protein - Chlamy
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W			
	1:	x{20}(L)x(T)(A)xx(A)(T)(H)xx(P)xxx(G)x{57}			
		MKGPAKSLALATLATAFSATHGPIAYIGICQTGCNALAVACYAAGTFGVPWGAAIPA			
1	S25969	ck: 7679	len: 69	1	hypothetical protein 69 - liverwort (Mar
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W			
	1:	x{131}(P)x(P)(A)xx(L)(T)(R)xx(A)xxx(A)x{142}			
		MWATKRSATTIKERELKRNRIYVHGRLAORPTPLNATITLYLMPFGFILSIIRYFNLPUPER			
		LFAELSDRIIVPVAAMCCKOGMENGTTVAGVKEWMDPEFFFNPRBSYEATFELRLPEE			
		MTVNGSGKTPIEVANYVOKVIGAVLGECETELTKRKYLLGLGNDGKVESINNTKK			
1	T01531	ck: 1750	len: 289	1	hypothetical protein A_IG005110.4 - Arab
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W			
	1:	x{131}(P)x(P)(A)xx(L)(T)(R)xx(A)xxx(A)x{142}			
		MWATKRSATTIKERELKRNRIYVHGRLAORPTPLNATITLYLMPFGFILSIIRYFNLPUPER			
		LFAELSDRIIVPVAAMCCKOGMENGTTVAGVKEWMDPEFFFNPRBSYEATFELRLPEE			
		MTVNGSGKTPIEVANYVOKVIGAVLGECETELTKRKYLLGLGNDGKVESINNTKK			
1	T05405	ck: 4069	len: 222	1	hypothetical protein F10M6.130 - Arabido
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W			
	1:	x{170}(L)x(P)(L)xx(L)(S)(H)xx(V)xxx(G)x{36}			
		MGIQSVLEPASYERRETELLIMRBEQDDESHWCVCEBFLICVAMDELOLGGSGSYTAGRC			
		SHEIVRRRGSTIEIAGFSRVRVKGKLPFKDKMLKKNNSNDLDLG			
1	T00559	ck: 9966	len: 296	1	hypothetical protein F12L6.15 - Arabidop
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W			
	1:	x{144}(L)x(S)(V)xx(L)(T)(H)xx(V)xxx(V)x{136}			
		MEHFRPDILSCPLLELLIYITISFPEBSARLPLVSTRFSVMQALILVAHSHNGSIEDISHA			
		TSLTIDSPKTLHLISGCPKRLRLYLEIISFKLFPHOGSLIFIKIHEHNLKRAV			
		FVTOGSPSYNNALDIDGLSLTIRKNSQSLTICRMWPEVTRLVSKFCSKIRLKYVITFVNKR			
1	T01617	ck: 1414	len: 392	1	hypothetical protein F19F24.9 - Arabidop
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W			
	1:	x{116}(L)x(T)(I)xx(V)(S)(H)xx(V)xxx(L)x{260}			
		MVIRSNLSKRRFSFKRRSRFSDSSENSPLAASSTCVGSGFQTFDQFQPKKCSFQGIYD			



```

109662  ck: 3889  len: 101  i multimeric endopeptidase complex chain
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(177)(V)X(S)(P)X(L)(T)(K)xx(V)xxx(V)x(8)
1:  CDPSONYGMKAGALIGANNOAAGSLKODYKDDITREAVNALAKVLSKIMDRSLSDLELA
S23774  ck: 5889  len: 402  i tltose phosphate/3-phosphoglycerate/phosphat
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(83)(V)X(P)(V)X(F)(S)(R)xx(A)xxx(G)x(301)
1:  MESRVSRATLTSPLPLNLRPLANALSPKSGSDGNLWGRQLRPELCSVLKKGAS
TSNVSEFAAVASFTIRKALEPFEFNAASQFLGSLPITLMLAPVAVIGVSMAS
LTELSPNMLGFTISAMISNISTFYRSIKKATDMDSNTIYVSIIALVICPALIEGPTLKTGFNDALAKVGLKVFSDLEFW
S22500  ck: 9347  len: 256  i embryonic protein ecp31 - carrot
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(185)(A)X(T)(M)xx(A)(S)(K)xx(L)xxx(L)x(155)
1:  MSQOQPRRPOEOPIRYGDVDFVSGOLSSOPVAPNDASAMQAENNVLTGTOKGCPSVMOGASAAA
AOSADYANARTMNVASKTRIGDVLADRSIRLAEDEKAVTREDAGVYGAVERNNPEN
MTYPCGVASMAAARLNDDPTF
T16984  ck: 4925  len: 165  i transcription factor homolog BTF3 - curled-
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(108)(L)X(S)(L)xx(A)(S)(R)xx(P)xxx(A)x(41)
1:  MNVEXLRKMGAVRITGCKGMRRKKAHVKTITTDKRLQSLTKRIGVNAIRAEVNIFFKEDVYI
S47086  ck: 1305  len: 263  i p17a protein - rice
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(32)(V)X(S)(L)xx(Y)(S)(R)xx(L)xxx(A)x(195)
1:  MEDGGHFEVHGLGAGCWMVIRVVALRAAGRAVALDMAAAGAHAPADAVGSLSEYSRPLDN
RPGNYIIDDPLMKDDTLTLEGNYGSGVSKRVFVPELTLASLKDALDLYRTRARAGDRLIDALA
BELAGADHMACSKRELCILLRIAAKYD
S14884  ck: 7397  len: 146  i hypothetical protein 7 - yeast (Hansenula B
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(44)(L)X(S)(L)xx(Y)(T)(R)xx(A)xxx(T)x(86)
1:  MGSTSGGLYSIFISALAKSLKRELQOOGGEVPELTLASLKDALDLYRTRARAGDRLIDALA
S43448  ck: 5386  len: 215  i FUN54 protein precursor - yeast (Saccharomy
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(33)(A)X(S)(L)xx(F)(S)(K)xx(L)xxx(M)x(167)
1:  MKRSTALSPFVTLALVNVSAASSYAPVAISLPASKECIYDMVTEDDSLAVGQVLTGSL
ESRLTWLSTLIITIIIVASIAQVLLIQFLTFGRKQNYV
S19425  ck: 1969  len: 317  i hypothetical protein YCR015c - yeast (Sacch
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(93)(L)X(S)(V)xx(I)(T)(K)xx(1)xxx(1)x(208)
1:  MKTIIISDDEITTRVDTICTIAKLEPYLNLPRKPEMHFTYMDGYHKRYNGTRSLPLSSGV
LKRVSDKQSQSYNGEFDRLTLTGSDKRVILIEIDKIDSGCNKSGNSYVIGDS
ETDLISILHPSITNGVLLINPOENPSKFIETEKIIGPKDKISFEADNGPAMVLFCEKEGKAYLVKSVDSLKDLMQVTKM
S70131  ck: 1796  len: 365  i hypothetical protein YDR273w - yeast (Sacch
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(137)(P)X(S)(V)xx(Y)(S)(K)xx(V)xxx(L)x(212)
1:  MGRKNRKGEMNNAKTSFLKVENIKNTNGLSLPDSODYTNVEKESSPRTDPLTLKHEVNTKIDS
KIESELISLAHIGNISIPGNSNRNREVAETRVAAEGNLTISNDSTLKEKRG
EERKEKGEKGVNSLKGAAVYVAKSLKNNRIPYTVKRNESNNLFDVLNCDSEDEEOVETINTSNOEKKNNGGNTVEPEAORSD
1:  NCKNKLRRFYNNVYTSWSPDEKADAMVMAKRSKKEQEVYINEQALFSLSWMT

```

LINKGLSYNEIITNOIELEEDGHPDITLLEEFSSRLNIFPNTNSQIFQONNNNTLILPKDSSGQEFLLSQTLADALTSFHSN

1  
S67699 ck: 4697 len: 193 i probable membrane protein ydl151c - yeast  
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x{16}(L)X(S)(P)X(X)(T)(K)X(A)XXX(P)X(101)  
MSTSLFSLSPSSSSMRLNRSNFFLNFLLDLPDLPSSSSASFSLSLPSLIVSRPCTDND

1:  
RPFLLILLNECPA

1  
S60916 ck: 7226 len: 139 i probable membrane protein YN266v - yeast  
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x{82}(A)X(S)(L)XX(L)(S)(K)XX(I)XXX(S)X(41)  
x{12}(I)X(S)(L)XX(I)(C)(R)XX(I)XXX(L)X(129)  
MMLNHTYLLSLYFLRKASNFENSSSSSFCSFLVFLVFFVDFCSITSFLLISFGLISFLLF

1:  
S66918 ck: 8726 len: 157 i probable membrane protein YOR044w - yeast  
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x{12}(I)X(S)(L)XX(I)(C)(R)XX(I)XXX(L)X(129)  
MTEALEIVLLVIOSTLOXICRQICIAFLIPPLGLYAFDELFLYVYRMILYSQMFYKRLGRSKTN

1:  
JT0589 ck: 609 len: 331 i pectinesterase (EC 3.1.1.11) precursor - As  
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x{11}(A)X(I)(A)X(A)(S)(R)XX(A)XXX(A)X(304)  
MVKSTLASVLEFATLAAKSRMTABSGALVYAKSGDDITISAVALDSTSTGTQIFIEESYD

1:  
EGAVDEIFGCHAHARHECDIFVLEGGPSASITANGRSSESDSTYIHKSTAKM  
DGNVSSGTYILGRPMPSOYARVCFQKTSMTDVIHNLGTEWSTIPNTEVNTVEIGNTGCAEGRANFSELEPTITMILGSM

1  
S03833 ck: 3269 len: 319 i hypothetical protein 1 - chestnut blight fu  
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x{62}(L)X(T)(P)XX(V)(C)(K)XX(V)XXX(L)X(241)  
MAOLRRPSOLVSESVPTVDPFVSRTBEVVPAGCITLMEYRDSGDPVPSLHSGDLRLRTH

1:  
TTGMCVAVDYLRLQWVGRRSFGFQLEKRSADVHVVDVADYEOEDGALFYQ  
AIIGLAEKDPRIARIGRLNPLAFAFGSALRVEVPQVTRRGRSTRMTRGDPITVFMQWYGHQHPGSCSGVGEERQFRS

1  
A34051 ck: 7750 len: 201 i NADH dehydrogenase (ubiquinone) (EC 1.6.5.3  
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x{106}(L)X(S)(V)XX(A)(T)(H)XX(V)XXX(G)X(79)  
MASRVTGVTAGTAGGVSVQKITVQSVGWERTIRAFALDPRNSGVPLVPRNRPSPGSLDP

1:  
WDYKTIQLAEEPSYSEAYFCRSFS

1  
S16556 ck: 4238 len: 289 i hypothetical protein 1 - Podospora anserina  
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x{1172}(A)X(P)(I)XX(V)(C)(H)XX(A)XXX(I)X(101)  
x{1172}(A)X(P)(I)XX(V)(C)(H)XX(A)XXX(I)X(101)  
MSEPFVAGLAPGAGLFNRCVACFEYVQUGRFGEDYERDCLRLDAKARLSWGAAYVINDPFR

1:  
AVCHKLAEIEIEVEDEASLTLLIKDAGGIDAMDAOKIDAIVGRNSAKDIRT  
EERARVQLGWVVAALHGRISDQITNSVETVYVKGESVLLIGNEXGKGFWND

1  
S16557 ck: 4187 len: 289 i hypothetical protein 2 - Podospora anserina  
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x{1722}(A)X(P)(I)XX(V)(C)(H)XX(A)XXX(I)X(101)  
MSEPFVAGLAPGAGLFNRCVACFEYVQUGRFGEDYERDCLRLDAKARLSWGAAYVINDPFR

1:  
AVCHKLAEIEIEVEDEASLTLLIKDAGGIDAMDAOKIDAIVGRNSAKDIRT  
EERARVQLGWVVAALHGRISDQITNSVETVYVKGESVLLIGNEXGKGFWND

1  
JN0451 ck: 2342 len: 312 i phosphoribosylanthranilate isomerase (EC 5.  
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x{62}(L)X(P)(L)XX(V)(S)(K)XX(P)XXX(I)X(234)

1:  
GSTTSSVNAALNGRDVYLALCSISSEVEVEYVEYKVCYLVEALMRASDTAKFLRSILGL  
GIDLVLHGDPEQMANFIPVPAVYKVRVSPGEGVIRRGNGNLTLDADAGA  
SGSGGEGKAFPMERAKRLIOSGEGSBGHPVLLIAGLTPENVALNRLVAFGVMNSAVGSKREGRSRLRLRS

1:  
B44418 ck: 3830 len: 412 i surface antigen - Trypanosoma brucei  
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
x{112}(I)X(T)(V)XX(P)(T)(R)XX(P)XXX(T)X(284)  
x{112}(I)X(T)(V)XX(P)(T)(R)XX(P)XXX(T)X(284)  
MCIEOLVSDDEEFKFAVACICLIGLGVLLIGVFLFSDPDNKRKAVSAFDPRLSEW

1:  
EKGCYCKQEVYLSKLYVAASDQKGEYREESTOYASLSEFGLHSOGYEYVQDKVO  
VOLYSEGPFLALBEEMGESEFVPRRTGICVAGSILLILEIACVAVCECLKRGSSNDTGDPTDQDGSPTTYGQS

1  
T14622 ck: 9305 len: 425 i hypothetical protein - Trypanosoma cruzi  
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
x{143}(L)X(T)(L)XX(L)(T)(R)XX(M)XXX(M)X(1266)  
MRLPRLLESLINOVGILNRSPPSSGKRGSGRISTDARLRLKRAVIEDALSTIVG

1:  
VHRLCAAPLYPFIYVIGNITITSSKSGATLMEKQVLRNDPCHASMGCEERSGAT  
QYTFIWMVRDHTHRAVSLDKFRVSVKAPMLNSITIAEVEVASFLIAAALITGTSQCDYFFIKAVRRRLSALNNGIYQETSP

1  
T15458 ck: 2648 len: 331 i hypothetical protein C08A9.7 - Caenorhab  
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
x{102}(L)X(S)(V)XX(L)(T)(K)XX(I)XXX(L)X(213)  
MVSATRVPRRSSTTSATVAAQRTSPSLMPASFPITMDEVLEKREFEVNVASKDIAMKALIT

1:  
FDGLFEVENNEVHNLPMMEKRVESANLHTEQANNEFDIOTENNMNMGATDSF  
SQIRHETSSPMNAQSIASIEYENVPRQIDTADNIGOVKOLVDPHDBRANFEVLEKTVLELRDPFTNAGVFEDMSL

1  
T15482 ck: 121 len: 217 i hypothetical protein C10A4.3 - Caenorhab  
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
x{96}(L)X(S)(L)XX(P)(C)(K)XX(L)XXX(M)X(105)  
MNTSRSPILLMOCSFISICITITANTESINHRILIPRQCLTGNPHQEPYQILNHLISOI

1:  
IEVRYDQPLPWTEQDLDAVSYSLFDFSDQLKREDC

1  
T15182 ck: 9118 len: 150 i hypothetical protein C18E3.4 - Caenorhab  
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
x{78}(I)X(T)(L)XX(L)(S)(R)XX(L)XXX(L)X(56)  
MSLTMDKCSRMKRYITVATTITTLDELITVIGACVLSCTVSVSICNQGFVVDIKALLXI

1:  
T15715 ck: 4709 len: 280 i hypothetical protein C30G12.2 - Caenorha  
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
x{197}(M)X(T)(A)XX(M)(S)(R)XX(L)XXX(A)X(67)  
MADVADPEVESRORPKNIMITGANGRIGLVKHFLEVDGIELLIATCRNEKADDELIALK

1:  
GCDFGPMKFEHYHFLVNSOEMTATAYRMSKALNFAROSMELESKYHILVTSFCG  
MVRTDMDGNADLDVNESTKTLISANILRLNRRNGLYFRELHPIPN

1  
T15729 ck: 770 len: 304 i hypothetical protein C31H1.6 - Caenorhab  
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
x{155}(A)X(T)(M)XX(V)(T)(K)XX(L)XXX(I)X(123)  
x{155}(A)X(T)(M)XX(V)(T)(K)XX(L)XXX(I)X(123)  
MSDBSCISPOKSPAESPEBOEKMKSNRNEANOSFYKGDADAKSDLLIDLEVYNNMAKIMEDVT

1:  
NHIGTSEELTWEVGDSCVYKPEYVYAEVTEEDVTEPRINRSKYLIVT  
ITNLEKNSLVITIEGICDILRGMDYIAKFRFDGPAIYTERGCIIRFGLEKVRINRESLNSRXY

1  
T16486 ck: 1289 len: 374 i hypothetical protein F56D3.1 - Caenorhab  
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)  
x{144}(L)X(P)(A)XX(V)(T)(R)XX(I)XXX(V)X(215)  
x{144}(L)X(P)(A)XX(V)(T)(R)XX(I)XXX(V)X(215)  
MKRLILAFMLGVVAKQOEKVEALSAEADONKERTKIIVGILIDGVRLEFVNDGGIYV

1:  
APPRVIAPOOPQFRIQICPQOPFPQAPAFQAPFTAPFQAPFT  
RPPRPQPNFATRPACOPAPFTPAPRFPQSGONLRSGOASIFISTPISGPSRLFTFRAVAVYDQACARTCHEPNCASN

T16528	ck: 3866	len: 419	1	hypothetical protein K02F3.4 - Caenorhabditis
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{100}(V)X(S)(V)XX(W)(T)(R)XX(M)XX(S)X{303} MCAQVPSLLPFWMSSTPLATSEDTLVGCGRTTLVLFRKHLPSILLRHVSQRPPSSVS ECSLPQTHOBNCDLIYVLDQCTTTGGSKRRSDRVEHKAASOGGDTRVGTASNGEAKREKKMSRPLGFT EFVDPEITSFDAEDFPVFEAPPAANSSAASPADFLDLSKSCDITSIDATIHHDSIASPSNASHISPRASDDIPDEFFPOL			
S40941	ck: 8579	len: 219	1	hypothetical protein ZK632.9 - Caenorhabditis
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{146}(A)X(T)(V)XX(W)(S)(H)XX(A)XX(A)X{x(57) MAAAEEGSGDLDTTGSSKRKSBRVHKHAASOGGDTRVGTASNGEAKREKKMSRPLGFT EENAROCNRKGLNLVKPTGSASPRYLQTPPKNAVAETTESO			
S44857	ck: 5825	len: 170	1	PAR.1 protein - Caenorhabditis elegans
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{1}(L)X(T)(I)XX(L)(S)(R)XX(M)XXX(L)X{309} MTLTFEVBASTLKTKTLLEMRIGLTSLFTIGLNCESLSGAIQRLPNESKLSPEFK VHRPIEKRCPODLQTLQMLITSEADQVLSKEPHPLAYLVEVNEVFGEERVLVPDHTIFLDSPFERHPRITREGGGGCPFX			
S44865	ck: 399	len: 336	1	R05D3.9 protein - Caenorhabditis elegans
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{11}(L)X(T)(I)XX(L)(S)(R)XX(M)XXX(L)X{309} MTLTFEVBASTLKTKTLLEMRIGLTSLFTIGLNCESLSGAIQRLPNESKLSPEFK VHRPIEKRCPODLQTLQMLITSEADQVLSKEPHPLAYLVEVNEVFGEERVLVPDHTIFLDSPFERHPRITREGGGGCPFX			
S06613	ck: 4664	len: 196	1	chorion protein s19 - fruit fly (Drosophila
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{160}(I)X(S)(P)XX(Y)(S)(K)XX(L)XXX(A)X{x(20) NMTEFALIFTSACLAVGSCGGISPGITGGPINGLRVSYISGQSOGSQAAMAAASAGDNGPY			
S69241	ck: 7903	len: 298	1	Dreg-5 protein - fruit fly (Drosophila
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{131}(A)X(S)(L)XX(L)(T)(R)XX(L)XXX(A)X{x(151) MTTAARKVLACCLCAAFIOISSSAIPIMEFLTRNEMSHLSTFAQLVSVCHCKSTAIVGGLPVN DKAVGYGPOSSELPAVALITSEPSKREFLGPVIIRVRPSPVEDKMPFIDE DLPLYLSWSGSRRAQOPDRHNOLIKAAALRLHPARERPATPDPAEASVPATGVRSRSEDPOA			
A24254	ck: 68	len: 255	1	ecdysone-induced protein - fruit fly (Drosophila
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{51}(A)X(S)(L)XX(L)(T)(R)XX(L)XXX(V)X{188) MSLITSVTHPELKDLSLVNEOKELNISVHVANTKATATFGCMCCFAGASYLGATRGVLRIT GHKDLASSLNTSPILTOISYAVALTRLNGYLAGVGIEBQFKAEAETTSDAHPAVLL LPRGERRPGSLLLTMVVRHR			
S78042	ck: 1000	len: 351	1	Ig mu chain C region, membrane-bound form -
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{101}(L)X(T)(L)XX(L)(T)(R)XX(A)XX(G)X{315) VOSAKRSLEFNWQCSASADGLVLCVCRDASADGSLFKNDSSGALLDVQYRAYVQATGTYTS EWTSSTGYVCKEFOOKNNHNFKEKSIAPGDTKQVQITPSTEIDLTIKRGQLEC RAEGDTGKSTIKMLIGNREJLSSNLSSTVLSQTHIGEWEINGEIFCEVHEAEFTQQIEKYTKFKENVCIMSTELFHXYEMDCI			
S69288	ck: 5524	len: 102	1	early lactation protein precursor - brush-b
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{28}(A)X(S)(L)XX(L)(S)(R)XX(P)XXX(L)X{58) ELRREQAREBEQGARIQALSAVAEHAARIALQADNOALDIAEADPPSP			
T14143	ck: 3550	len: 68	1	Arpase subunit 8 - orangutan mitochondri
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{34}(P)X(T)(P)XX(F)(T)(K)XX(P)XXX(P)X{18) MQUNTWLTVIIPPTLALFLIQLKLNLSHLHPRPFRFTKTHAKRWELKWKTIYSP			
JC5761	ck: 447	len: 225	1	cytokine-inducible SH2 protein 3 - human
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{185}(P)X(T)(L)XX(L)(C)(R)XX(V)XXX(L)X{28) MYTSKRFPAAGSGDPLDTSIRLTSLRKTSKSEIOLVANAKLDOESGFYMSAYTGGANLLISAIE LSRPLSSNVATLOHLCRKTYNGHDSIERVTQLPGRIRFDQIDAPL			
S43928	ck: 3434	len: 310	1	estradiol 1beta-dehydrogenase (EC 1.1.1
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{193}(L)X(S)(M)XX(A)(S)(K)XX(V)XXX(S)X{101) MGDLDEPFLTLTDLCLACLAFCVGLTNLWNKWLPKFSLMCOMAWAITGAGDIGIKA GLULINSSGIALPWPPLYSMSYSAKFAVCFARSAKAEVYIOLITFYAAVST AMTYLNTVITTADEFEVESINVTYTIGTCGCLAHEHILANGSELIPAMAIFYSGAFORLLTHYVAYLKNTKVR			
JE0175	ck: 3945	len: 317	1	frazzled protein-1b - human
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{117}(P)X(T)(A)XX(V)(T)(K)XX(L)XXX(L)X{131) MRAAAAGVRYTALAILGLALMAARCEEHVYQMOAPELHGSRYSYSPQCCLDIPADLPIC TKIACGEMHSADGMEOMCSSDFEVYAKRIEKNDRIGAKRKKLLKG PLKRDKRLRYLHMKNKAGGCPQDLSLAVGNKRDGGLLMNAVYRMDNKKREKFVAFREMFSPCLSYPPFYGAEPH			
I77373	ck: 9167	len: 22	1	gene N-ras protein - human (fragment)
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{4}(L)X(T)(L)XX(L)(T)(R)XX(P)XXX(C)X{2) ETCLDTELRLRLPLCTHQ			
T14788	ck: 6935	len: 112	1	hypothetical protein DKFpZ564A122.1 - hu
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{74}(V)X(S)(L)XX(L)(C)(R)XX(I)XXX(M)X{22) MTGGDYLYLAMDMCMLALMIKWRFERWRIDEPYPOVAPVNSGCLCAPSTDGSVLAAGT			
T08722	ck: 6844	len: 381	1	hypothetical protein DKFpZ566D133.1 - hu
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{102}(P)X(S)(A)XX(Y)(C)(R)XX(V)XXX(C)X{263) RMTQAFREPELPTWTOTIDEDBALDELQYAGVSPCLORTGISLINIPAQVCIASEKOE QAYDEPAGWGMEVNLNHGCFGAGLOKNSKSGLTYPATSTVATIHMSRIMS DSDDLTKRLRLHGNDEVHIWMSEHTRDYGRIIPREGDVLIVYPMKNHMFESIOMRKEVEVFFGLDGALVNGKVLPIVVR			
T00702	ck: 4343	len: 196	1	hypothetical protein F25965_1 - human (f
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{118}(P)X(P)(P)XX(A)(C)(R)XX(I)XXX(L)X{162) SWRSKRGKTMMWLSCSOWRMYWIGSVVMKRLGKRIKPOTINTYIKRLFDSVDLAGESENT NQLFYSKMILNEMERO			
A44478	ck: 3379	len: 373	1	probable cell growth or differentiation
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{129}(P)X(P)(A)XX(L)(S)(R)XX(P)XXX(V)X{128) MLUDEALMKLVWDGIGORVAVGSCDTGOEVVILAOLAIGQTGHFVLVQRLEREKROLLPO ELRREQAREBEQGARIQALSAVAEHAARIALQADNOALDIAEADPPSP			



MASATERLHODLAVOEROSAEVOSGLALVSALEAERLQAQAELEELNRELROCNLOQFIQOTGALPPPPRDRGPGTGGPLH  
 1 B44478 len: 337 1 probable cell growth or differentiation reg  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{129}(I)x(S)(I)xx(L)(S)(R)xx(P)xxx(V)x{192}  
 1: MLIQANAEELKAWDGIORVVGCGVECTCEVIALQAIGTGRVQLREKERQGLQPOECH  
 ELREOANEREGQARLQALSMATAEHAARLQALDAQALAELEALAEAPGPPSP  
 MASATERLHODLAVOEROSAEVOSGLALVSALEAERLQAQAELEELNRELROCNLOQFIQOTGALPPPPRDRGPGTGGPLH  
 A61188 len: 358 1 probable transcription factor SCI - human  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{175}(L)x(S)(I)xx(L)(S)(K)xx(P)xxx(L)x{167}  
 1: MLCFOLIRIGGREGDXTTHPPRAGGCTTRDLGHRADCDVALRPOQEBGLISGHAELHAEPK  
 SIGSLSKRPOPLFTFTSGMGPKSLTPAPRPGVGTGPPAPRPNRRKRVHRYALAE  
 LDDESEPPENPPILMEPRKRLNDRKAPLPTGRRRORPKRTIVSAPMAPVAGTQQLVACPRRQMPFSSVNAVTSQMWPLY  
 S63679 len: 65 1 signal transducer and activator of transcrip  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{129}(I)x(S)(V)xx(V)(T)(K)xx(L)xxx(P)x{20}  
 1: GTFLLRFSESSGEGITCSWHEHODDDKVLIVSVQPYKREVLQSLPLEIIRHYQLTEENIPE  
 A41795 len: 391 1 somatostatin receptor 1 - human  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{188}(I)x(P)(I)xx(F)(S)(R)xx(A)xxx(G)x{187}  
 1: MFPNGTASSPSSSPSSPSSGEGGCGGAGADGMEPGRNNSQNGTLESGGSAALLISFIS  
 NLGVWVLSLVLPIPIVFSRTAANSDDGTACNMIMPEPAORMLVGFVLYTFIMGFL  
 LPVGAICLCYVLLIAKMRVVALKAGMOQRKRSERKILTMVMVMVFVLCWMPFVYQLVNFAEODDATVQSLSVILGYNANSCA  
 JCS627 len: 225 1 STAT induced STAT inhibitor 3 - human  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{185}(V)x(T)(L)xx(L)(C)(R)xx(V)xxx(L)x{24}  
 1: MWTSSKRPAGMSRPLDTLSRLTFSKSTYQLVNAVKRLDQSGFYSAVTGGEANLLISAEPAG  
 LSRLPSLNVATLQHLCKRTYNGHLSDEYKVTQPGIRREFLODTNDFL  
 S41063 len: 183 1 translocon-associated protein beta chain -  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{182}(I)x(P)(A)xx(V)(S)(H)xx(V)xxx(L)x{85}  
 1: MRLTSFYVALFAVTAQEAEGARLLASKSLINRYAVEGRDLTIQYNIYVNGSSAALDVELSDSPFH  
 PKTKKN  
 A36679 len: 183 1 signal sequence receptor beta chain precurs  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{182}(I)x(P)(A)xx(V)(S)(H)xx(V)xxx(L)x{85}  
 1: MRLTSFYVALFAVTAQEAEGARLLASKSLINRYAVEGRDLTIQYNIYVNGSSAALDVELSDSPFH  
 PKSKKN  
 A56446 len: 268 1 Ig heavy chain V region (3H-3H scFv) - mous  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{152}(I)x(P)(A)xx(V)(S)(H)xx(V)xxx(L)x{85}  
 1: MAOVKRLDESARELVAGASKCTTSFNKIDTQDMHWKORPEQGLWIGRIAPANGITKYDPKH  
 PKLMVYVTSHTLPBGVAVARSSGSGSSGSLTSSSEGDAATVYCQOFTSSPFTFG  
 SGTKLEIKKSAHHHHHGAADCKLISEEDNGAA  
 B61188 len: 265 1 SCI protein - mouse (fragment)  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{173}(L)x(S)(M)xx(L)(S)(K)xx(A)xxx(T)x{76}  
 1: MLCFOLIRIGGREGDXTTHPPRAGGCTTRDLGHRADCDVALRPOQEBGLISGHAELHAEPK  
 GSLSKLQAOPLTFRCGREGKRLALPSQGEAAQVSPAPRPNRRKRVHRYALAE  
 DDEVPSPSLVTEPRKRLNDRKAPLPTGRRRORPKRTIVSAPMAPVAGTQQLVACPRRQMPFSSVNAVTSQMWPLY  
 S63681 len: 64 1 signal transducer and activator of trans  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{129}(I)x(S)(V)xx(V)(T)(K)xx(L)xxx(P)x{20}  
 1: GTFLLRFSESSGEGITCSWHEHODDDKVLIVSVQPYKREVLQSLPLEIIRHYQLTEENIPE  
 C41795 len: 391 1 somatostatin receptor 1 - mouse  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{188}(I)x(P)(I)xx(F)(S)(R)xx(A)xxx(G)x{187}  
 1: MFPNGTASSPSSSPSSPSSGEGGCGGAGADGMEPGRNNSQNGTLESGGSAALLISFIS  
 NLGVWVLSLVLPIPIVFSRTAANSDDGTACNMIMPEPAORMLVGFVLYTFIMGFL  
 LPVGAICLCYVLLIAKMRVVALKAGMOQRKRSERKILTMVMVMVFVLCWMPFVYQLVNFAEODDATVQSLSVILGYNANSCA  
 S66270 len: 457 1 kynurenine--oxoglutarate transaminase (E  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{195}(L)x(P)(A)xx(A)(S)(K)xx(L)xxx(L)x{246}  
 1: MFRSAALSVHLMPLMKRNRKAGASCTTRCTHOSLIMTRKLRDQRRLDGLDQMLWEPFKLRETD  
 LKSPAPRKGKLGASNDMDLDAPELAKSKTPRTKILVITNNPNNKRGYFSEMELEI  
 ANLCQHDVYCSIDDEVYQMLVYDGHQHSINSLGMDMDRLITLIGSAGKSFATIGMVGWGEEDNIMKHLRTVHONSIFHCPTQA  
 S41286 len: 223 1 latexin - rat  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{34}(V)x(T)(V)xx(A)(S)(K)xx(L)xxx(G)x{173}  
 1: MEPIIPFAASRAASVADENQINQOOGIPNKVFEVQVQOAKSEDIPGRGHKHLKFVSVEEII  
 ELIDYVLLHDVASQELIIPWQVLMHPQYGVKYNRLPREKAPAE  
 A39297 len: 391 1 somatostatin receptor - rat  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{188}(I)x(P)(I)xx(F)(S)(R)xx(A)xxx(G)x{187}  
 1: MFPNGTASSPSSSPSSPSSGEGGCGGAGADGMEPGRNNSQNGTLESGGSAALLISFIS  
 NLGVWVLSLVLPIPIVFSRTAANSDDGTACNMIMPEPAORMLVGFVLYTFIMGFL  
 LPVGAICLCYVLLIAKMRVVALKAGMOQRKRSERKILTMVMVMVFVLCWMPFVYQLVNFAEODDATVQSLSVILGYNANSCA  
 JQ4294 len: 274 1 ammonia monooxygenase (EC 1.-.-.-) A cha  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{133}(M)x(T)(V)xx(L)(T)(R)xx(M)xxx(L)x{125}  
 1: MSRTDEILHAKMFPESKMSRMDAIFPLICILLVGTYHMHFLLAGDDWDFLWDKDRQW  
 GVLLSLADYTGFLVVRGTPEPYVALILOGSLKRTFGGFTVLAGFSAFVSMILFCV  
 WMYFQKLCYVLLIAKMRVVALKAGMOQRKRSERKILTMVMVMVFVLCWMPFVYQLVNFAEODDATVQSLSVILGYNANSCA  
 C75042 len: 308 1 aspartate carbamoyltransferase, catalyti  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{17}(V)x(S)(I)xx(F)(S)(K)xx(L)xxx(L)x{285}  
 1: MDWKRGVDSISIRDSKSDIEVIALTAERLEREKQGELEAKGKILATLFPPESTRRLSPE  
 TTYDELVLLISPELLRPHRIIVELBEKRGKAVETTLTDVGLKDVLTAVTRIOKE  
 RFPDEOEVLKAGSYOVNLVLEKADDELRIHMLPRVDEIHPEVDNTKHAIFYQVNGVAVRALLATLVGLVI  
 D71333 len: 477 1 conserved hypothetical protein TP0373 -  
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
 x{174}(V)x(P)(L)xx(V)(S)(R)xx(V)xxx(L)x{281}  
 1: MSESROKIHPLVAVASRGHPLPAPKSGTIVLAVSGSDSLALILAAHELAAPDFGVCACAVT  
 LLKVSRCVDEDTOTRGVRRERDASNMCKRYVNRHRLTALNAVLAAGRSGSD  
 KTFGISAHSFCAALITNRRECSHAMBEVPAALGTRILMRSPDLAEFLILRFLLOEACVRLGVSHRVPGALERCAREDDGV

1	G71287	ck: 2410	len: 365	1	conserved hypothetical protein TP0730 - syf
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(175)(L)x(T)(L)xx(A)(T)(H)xx(L)xxx(X)(174) MARASSLSLMSHSPHSIPSAHCAALRRKTCQSGSDASFQKAHEKSTLPKAPARFPYVYF TLNTATHTILIRASAPIVGFILKFKHALGYSYFVAGISYFLTFDFIDGKIDARAR RETNRVCELTADSDYALIGLISALYOGSVPLMFVILITRLSLQIVACVYALFGHPMTGSTAGKATVATMLTYLTLEARLLI
1	J00857	ck: 6140	len: 169	1	hypothetical 18.1k protein - Escherichia cc
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(124)(I)x(S)(I)xx(V)(T)(R)xx(A)xxx(V)x(129) MFDYQSKHPHDEACRAFLRNHVLQALERAGMNVOLIRNKNLNSQPHLLTAEIMLTDLEDS
1	T10288	ck: 562	len: 298	1	hypothetical protein 19 - Oryza pseudocausg
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(46)(A)x(S)(I)xx(L)(S)(H)xx(A)xxx(G)x(236) MLVOVNFIDLVLAHALFEGCSFAFYFALMAYVATIAFLELEDSANSHLAAFLGPIYI VRHDEPTVNONOGALTMALIALIYAHCHNMKVQKRETRKRNRANTITLMEI EKEDYAEHEEHSNRSWILSARRARNRPHNPFECOTOTONVSRLLIQRARRAOPYPTAE
1	G72757	ck: 9964	len: 175	1	hypothetical protein APE0052 - Aeropyrum pe
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(37)(L)x(S)(V)xx(A)(S)(R)xx(V)xxx(V)x(122) MPSALRTIAIAVNLVAFIAGAVIYSEYSDVSTLOSVTASRVTVQAGVAYLGYGATAYIG
1	C72759	ck: 3453	len: 114	1	hypothetical protein APE0064 - Aeropyrum pe
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(9)(A)x(P)(I)xx(V)(T)(R)xx(A)xxx(A)x(89) MTSLGACSIASPIILWTRFSTAPAKRPAASSIRAASSGVCVSSSTALEPRARLITSS
1	A72773	ck: 7998	len: 129	1	hypothetical protein APE0172 - Aeropyrum pe
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(82)(L)x(S)(A)xx(L)(T)(R)xx(L)xxx(T)x(31) MSPULISILINGADARVSSISLSTGFANFRVWICOMNVGAIANAAYMSSQGSTVLTSTAMWLG
1	C72775	ck: 1302	len: 100	1	hypothetical protein APE0189 - Aeropyrum pe
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(24)(P)x(S)(I)xx(V)(S)(R)xx(P)xxx(S)x(60) MLFIRSRISDIIFLDKLAIVIIIPMSISVSSRSRPLTSTLIVIMMAASSVASCDFDILA
1	C72784	ck: 9325	len: 313	1	hypothetical protein APE0258 - Aeropyrum pe
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(152)(V)x(P)(A)xx(M)(T)(R)xx(A)xxx(I)x(145) MMLVGSLDVIVIEVSRALGKDFOGGLMKSGINNOAKAELISSHPILBAKRVTEVRESYKH ERDVEGVAAVREFLEKFEKARVGTIGTNSAGTIGVLENGENKSLSP VHAALVPIDKIVBSVMDAVNALVQAFAFGFMPYVILISITGSPATGIDIEIKVLGAHGREVHVLLDNGRMAKASHPF
1	H72721	ck: 4962	len: 168	1	hypothetical protein APE0313 - Aeropyrum pe
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(8)(L)x(P)(I)xx(V)(S)(H)xx(P)xxx(V)x(144) MVGSEIVLKPISGFVSHGLPDSVRENIGVGIFLEVPEIAPGLDGLAGFSHITVIAIYHKSXGR
1	H72731	ck: 2984	len: 172	1	hypothetical protein APE0393 - Aeropyrum pe
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(65)(L)x(P)(I)xx(A)(S)(H)xx(I)xxx(G)x(91) MGSTLSSSIPSSVYLAISSAALAIIPSWARDISRAPLTKPENTRALLTWLGRSLPVA
1	G72736	ck: 2336	len: 180	1	hypothetical protein APE0429 - Aeropyrum
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(26)(I)x(S)(V)xx(V)(T)(R)xx(V)xxx(P)x(138) MSASVARRRRESKKEGEGEKSKDKLIISEVETTRDKVVKNPOMALYLLIDRLGPIHERTLQ
1	LGF				
1	D72737	ck: 7680	len: 120	1	hypothetical protein APE0434 - Aeropyrum
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(59)(P)x(P)(I)xx(L)(S)(R)xx(I)xxx(V)x(45) MRSPFMSYILFNLISGENLGGCGTATGPQRTGASLTSVYLPHLHFRSHCNLASPO
1	A72741	ck: 338	len: 160	1	hypothetical protein APE0458 - Aeropyrum
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(9)(A)x(T)(P)xx(I)(T)(R)xx(L)xxx(L)x(135) MVGGAULTAETPAPITRPLDLDSLTQASLMSGLVSSLTSTIPISPLPTMTPTIIS
1	G72640	ck: 9276	len: 147	1	hypothetical protein APE0559 - Aeropyrum
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(62)(L)x(P)(P)xx(F)(T)(H)xx(A)xxx(P)x(59) METTPYRAASTASTRTPGSHGASMTPTPLCLAMYSISLALAGSLSSWSLGTREPPIL
1	A72651	ck: 639	len: 119	1	hypothetical protein APE0638 - Aeropyrum
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(102)(L)x(P)(L)xx(A)(S)(R)xx(A)xxx(T)x(40) MMTRLDPPAGAGGIYYSRCLSATPTFSLAGFASARTSKALMGFWPPOLYTSTVANYXS
1	D72672	ck: 4161	len: 129	1	hypothetical protein APE0802 - Aeropyrum
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(102)(L)x(P)(L)xx(A)(S)(R)xx(A)xxx(T)x(40) MTPIKSLPSOLSCSLNSLDMNSFTRSPSSIPSPENGLIHGLVLTNPGLGLMLITIN
1	E72679	ck: 8681	len: 234	1	hypothetical protein APE0857 - Aeropyrum
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(173)(V)x(S)(L)xx(A)(S)(R)xx(P)xxx(G)x(45) MLLRNPKQGVETGPGGGODEGVVHPPIVRAPORLAKGAEMDLEVPDAGGAEIEMWH
1	G72685	ck: 9005	len: 228	1	hypothetical protein APE0902 - Aeropyrum
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(70)(A)x(S)(V)xx(L)(T)(R)xx(L)xxx(V)x(142) MEALYRSVEOLVNLMDLIDYLGSSDIPRVEPIYLEAKSKLSRPTAGIARKIIGQKHLEL
1	H72607	ck: 1372	len: 387	1	hypothetical protein APE1326 - Aeropyrum
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(165)(L)x(S)(V)xx(L)(S)(R)xx(L)xxx(G)x(1206) MAVSDSYSLVDVAVHKLRAELRGEDEBFKASRDMHARRPPICGWTHTITGCGCSTG
1					LEPGASPEERIEFMGRLVENAGLSATIFLPIVGTDEADIEDILSRARAKVNV VLGITRVAEGILRLRASGAVENGELIERLRLPWRKGEOLPIYSRDLKERISAKREMGFVYLPASCANVESHGCGCAACRUGP

1	E72626	ck: 6841	len: 146	1	hypothetical protein APE1469 - Aeropyrum pe
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{100}(L)x(T)(V)xx(L)(S)(R)xx(A)xxx(A)x{30} x{164}(R)x(S)(L)xx(L)(T)(R)xx(A)xxx(L)x{319} MKSAMDRFTSLVRFRTTNGEFLPAVLALALANLDRVFGNMRMLDEOMOCWRELOSS			
1	D72630	ck: 4651	len: 152	1	hypothetical protein APE1499 - Aeropyrum pe
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{88}(A)x(P)(P)xx(I)(T)(R)xx(P)xxx(T)x{48} MSLRSSRSKRSMSGSIITVLAIPLGLIGMSLCSIMSTLHLFSATISRTLSMPGMSGITPV			
1	H72538	ck: 9362	len: 175	1	hypothetical protein APE1597 - Aeropyrum pe
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{39}(V)x(P)(V)xx(A)(C)(R)xx(V)xxx(G)x{120} MDECRHLGVYIVYLANVYFVLAGSRHLHAHRRDRKPNVPTACRTVVTMLGAGTGDEVEYL			
0	D72545	ck: 2785	len: 194	1	hypothetical protein APE1648 - Aeropyrum pe
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{16}(A)x(S)(M)xx(L)(T)(R)xx(A)xxx(V)x{162} MMGRGLKMTLGDRLVAVSMGDLTKRAENAVNPANSIMMGGAAGALKRAGGSVIEEEMRAH			
1	ALGEGGECPADLRLV	1:			
1	E72557	ck: 8347	len: 132	1	hypothetical protein APE1745 - Aeropyrum pe
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{36}(A)x(T)(V)xx(E)(S)(R)xx(A)xxx(L)x{80} MRVLEPRGSFSLTASRVALLPESELILTSACRLNERATVYFSKAPAPRTLPGMTISPLACH			
1	H72557	ck: 8142	len: 181	1	hypothetical protein APE1748 - Aeropyrum pe
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{94}(A)x(S)(I)xx(I)(S)(R)xx(P)xxx(T)x{71} MYFGSVILGSPISVWLSCSRRAVLRSLILITLOSSSSSSHRKNSTRPRAIILITASAMPITF			
1	TPCL	1:			
1	D72577	ck: 2318	len: 100	1	hypothetical protein APE1900 - Aeropyrum pe
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{76}(P)x(S)(A)xx(A)(T)(R)xx(P)xxx(P)x{18} MIPVYVELNTSMRSEPKSQIAMEPVLRSILSGMLNREMDLITICSSIPSTLSFNNFSKLIS			
1	A72511	ck: 2835	len: 140	1	hypothetical protein APE2063 - Aeropyrum pe
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{66}(L)x(L)(L)x(L)(T)(R)xx(L)x{158} MLSSASAMSSDMPKTRLATASRSALITLPSLNSLPKAFSNTSAAGPLRSMTWLTFESHSPSS			
1	C72529	ck: 9995	len: 310	1	hypothetical protein APE2207 - Aeropyrum pe
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{45}(V)x(P)(V)xx(M)(S)(R)xx(L)xxx(A)x{249} MYRAEVRTESDIYDIAVYFMSLAVRLVTFELEENRKTAGIQVYFVMTMSRLILITSTAVFLG HVLDSPPSSEDFLIGYTAVKYSGSVIHLEIRINDLERLNDIRLALARMELY LEILVAVAVATIAVYIFFTISSVFHGNLGVASAOYASTOLILYSFLVPLTINMLLVDRARPEIQIKDTW			
1	F72471	ck: 430	len: 133	1	hypothetical protein APE2415 - Aeropyrum pe
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{82}(L)x(T)(I)xx(M)(T)(R)xx(M)xxx(S)x{35} MLVKRDISMLPRTIATAPIESRPNRATPVYAALTALEMLLAPRFPDTVAVAMLRKYADWTMO			
1	C70405	ck: 5299	len: 499	1	hypothetical protein aq_1220 - Aquifex a
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{1164}(R)x(S)(L)xx(L)(T)(R)xx(A)xxx(L)x{319} MKRFLVILVPLILYFNLGNVAMNPNESFVADSNAMKSGEFLTPYNGEIRLNKPPMT FVLGAVYFYLRLARRELKIKYAGTAVLFLSGWMPLOYLRHEEFLVFI KENKRIYALQBDPEYVALDINVSFLPYSFFFALFWALKERREISPLVWSPFLIFSIVMKIPYIMPAYMALITA			
1	D70454	ck: 3320	len: 269	1	hypothetical protein aq_1793 - Aquifex a
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{4}(M)x(S)(I)xx(Y)(T)(H)xx(M)xxx(A)x{249} MMSGMAIIMYTHFRMVKHALNWSSTLTONRFLTAELILFTLNLFLIPVANNVFAEFIN TEGTPRDFEFLITTAPEFSLKIKLAPSEVLFSSPVIGFSIFLFLQFWSHLF PPLFLITFVAFSNTLILITLFGVYSVSYLMKRECK			
1	A75196	ck: 70	len: 147	1	hypothetical protein PAB0052 - Pyrococcus
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{57}(M)x(P)(P)xx(A)(S)(R)xx(A)xxx(L)x{74} MSSLRLAMPPLNSISGATTLTSFVAIFKAEFLITNSLVLSGKSPSSSLKPTMTSPPR			
1	E75204	ck: 8207	len: 179	1	hypothetical protein PAB0107 - Pyrococcus
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{96}(M)x(T)(I)xx(A)(S)(H)xx(I)xxx(S)x{67} MDGYENKATNIKIRHIMKRDQDFLTSNPEISPLILVLIGSIFPDLDVFTFFSESLALH			
1	H75185	ck: 6767	len: 266	1	hypothetical protein PAB3346 - Pyrococcus
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{190}(I)x(P)(L)xx(I)(S)(R)xx(A)xxx(M)x{60} MTRVAPFDIDDMVSVFLDVAIVNTGPDSDVAKSMETSLKVOGNCILAFIDGKPYMGKICFLY VENEGGMYRGRITPPLVAISRKAIRKIMKATKELIAGRELIVPDANEDMDLIKRF SPVELTSCITRMRLGKVEENHVMFGILITAKG			
1	D71228	ck: 1741	len: 135	1	hypothetical protein PH0086 - Pyrococcus
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{42}(I)x(P)(P)xx(A)(S)(R)xx(A)xxx(L)x{77} MRGATTLTFLVAISKALTLTINSFVLSGKSPSSSLKPTTISPPRRSRREAIILKTKTFL			
1	D71110	ck: 6672	len: 102	1	hypothetical protein PH0651 - Pyrococcus
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{21}(P)x(P)(V)xx(A)(T)(R)xx(A)xxx(P)x{65} MLTILSLGNNIKAEFNVVYLPBPAPATRMFAGLTTPRSTRSORNAATSEFVYLVNFISI			
1	E71120	ck: 8797	len: 104	1	hypothetical protein PH0732 - Pyrococcus
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{42}(P)x(S)(I)xx(V)(S)(R)xx(L)x{46} MAHNHHYTPPKLHCHVAIPTEGKHRYIVSPSKDSCYQIPISIRVSKLKIISTYLVNSC			
1	H71124	ck: 5357	len: 272	1	hypothetical protein PH0767 - Pyrococcus
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{43}(I)x(X)(L)xx(A)(T)(R)xx(L)xxx(L)x{213} MANENRLLSGSSAVRTPGPRALVWALTEPLGKCCENALMLSSIVFATKTPCLCNAAYSI VTKAFKSTVFTIIPFOGPMILIPDMALISSSSSIFSPSPSPSPRPAVITITPL TFPSPHMISVSTNLGFTIIANSISSGSLTLALALP			

1 F71127 ck: 8012 len: 183 ! hypothetical protein PH0787 - Pyrococcus hb  
<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x[15](I)x(T)(L)xx(M)(S)(K)xx(L)xxx(G)x[152]  
MULLHSEGSVARDCTRTILIRMSKNRLPQGGKITTSKINISGAFSLVNLALIVDLAQPKYGLIAMV  
1: KNVDKV

1 F71039 ck: 7274 len: 447 ! hypothetical protein PH1606 - Pyrococcus hc  
<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x[186](I)x(T)(L)xx(L)(I)(K)xx(L)xxx(V)x[263]  
MKRIKSGIISAKITICMLIKKRIKINGRISLSDGLFPGDSHLSGKFLYINVEQLINVQDT  
1: EDALFGLVSIIMAKRLKLGKFSVEYLKIKKRIKIFEFLLADVIGLITKATITRFGRK  
YISSLYTKFSDIRKRLLYFLFLNLFPLMLDFIGMLSLPMLPECLLASRELYQWFGHYPMALVPLSFVASLEVAROLKRIYKIVK  
F71201 ck: 6730 len: 137 ! hypothetical protein PH1883 - Pyrococcus hc  
<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x[28](P)x(P)(L)xx(M)(S)(K)xx(L)xxx(P)x[93]  
MSAYTLTGLSLSVSONSESIPCGAVSMNPPEPLFMSKNOLVSLPSKNIGFTYSELVWYSSFPIS  
1: D71216 ck: 2646 len: 225 ! hypothetical protein PH1996 - Pyrococcus ha  
<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x[189](P)x(T)(A)xx(V)(T)(K)xx(L)xxx(P)x[20]  
MCSEFLMIADFEDGTAVDSQSCIEAFYALRTYTPWFGKRYMAKLTRIKELQFERPRFGHSGG  
1: GDENTEYVMWYFPYARFVTKELIDKIPHLHYIDMLSDLYRLSLV  
G75035 ck: 3225 len: 331 ! iron (III) abc transporter, permease protei  
<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x[37](L)x(S)(I)xx(Y)(T)(R)xx(L)xxx(P)x[278]  
MKRVTALIASSTTIVLAFSTGSVRIPISTILNISLSHSISLYTRGOLSGSPYIILGIRPRVMI  
1: WLFQGLAATATSKVPIMAVANGICGMGLFMSREPLILGEGSTALGDVNYTR  
LILRALILTGVAVSESSITIGFICLVSHIMKMFPGPHRRLLPSPALMGLILVLSPLISRTIYSPVEIPVGIVTALFGAPFAFYLLIL  
B75110 ck: 5732 len: 249 ! lipote-protein ligase A related PAB1916 -  
<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x[166](L)x(S)(V)xx(V)(S)(K)xx(L)xxx(G)x[67]  
MRPIPLIARPELOMADIDELILARSEGVAPDPIRYLVERPSSVTIGRFQSVRHVDIEKABELNLT  
1: LRDGIRSIYERAVTLTREGIKLSMQEYELLRNSEFRAFLPELEGTEYELELNO  
ELVEERYKDEDMENFK  
E49094 ck: 3704 len: 373 ! methylmalonyl-CoA decarboxylase (EC 4.1.1.4,  
<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x[144](P)x(S)(I)xx(A)(T)(K)xx(P)xxx(G)x[213]  
MEAFPAIDIOSVINDSGFLPRTGNALIMIVGLILTLARAREEPILLPIAFGCLLIAMPNGEE  
1: PPMVKLFTQKEREIEVQLREAVTREFKEIVYATIEISLPLPSTGLGMLMG  
NIFRESGVTDRISDTSQNALIMTVIPLATGGLTMSAHEHISLETIKIILGLFAFCGTAGVGLFGLKMSLVDSGKTNPLISAGAG  
T14222 ck: 7879 len: 344 ! NADH dehydrogenase subunit II - Xanthusia vlu  
<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x[137](P)x(T)(I)xx(M)(T)(H)xx(L)xxx(I)x[191]  
MSPIIOLISLSTAGTITMTTOSQSLIMVAGCEINTLAILPILSKPHPRATEAKTKFLQOTAA  
1: TLAASSTAHLMIVAIASMDSTITLITLITLISLPIISMNTHSSTMDMATY  
WTLSPTLTTPMLLILISLGGLPPLTGLFPLKMLIDLELTHHTPLATLALSTLSLEFYLRSLSTTTITILHAPNTIMSNNKRRKKYK  
S55472 ck: 8743 len: 299 ! pectin lyase (EC 4.2.2.10) - Mycosphaerella d  
<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C  
x[55](L)x(P)(P)xx(V)(T)(R)xx(L)xxx(L)x[228]

1: MFSTFVSLGLTAITAAPLAPGSMIDKVDKYRAASLEDVATARPRRCRSALPPSPRT  
HAADVIYTNFLHDHMKASLGSDSGADSKCHPLLVTVANNLNKNINSGRPA  
SQAQISTTTTRTCGMVSTPARVSRCLERTFTLSAPRLFRLLMDMLSLATTSVMARLLPRLV

D72485 ck: 7054 len: 242 | probable high-affinity branched-chain am  
  
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W  
x[179])(V)x(S)(T)(L)xx(I)(S)(K)xx(I)xxx(M)x(47)  
MEWTLTIIDRVYKRFGIIEALKGVSEFVRGDEVLIGPGACKTLTFNISIYMPDROR  
DLVTYLIEISKRGIIIMVEHMAYNNFAIRVYLHFGEKIAGCTPREVASNLV  
LEAWNGTGS

B71024 ck: 6905 len: 249 | probable lipocate protein ligase - Pyroco  
  
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W  
x[166])(L)x(S)(V)xx(V)(S)(K)xx(L)xxx(G)x(67)  
MRPIPLIVAPPVQNAADAILIASCEGPDTVLIVFSPSSVLTIGRFQSVRYDVNIERAKE  
LKDKGVKSIFERRVTLEREGIKVGRNEMEVLANSFRRAPLPDGEULTELELVE  
KLIEREYGNDRKNFOK

S72288 ck: 9584 len: 75 | ribosomal protein L23 - Plasmodium falci  
  
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W  
x[26])(T)x(S)(I)xx(F)(T)(K)xx(I)Vx(133)  
MKEVILNFYLYILFIYKYNLKNKFCLIIYSIKFYTKDIKKYIIKNFKIKLINNNKINI  
1: T10349 ck: 889 len: 374 | very-late factor 1 protein - Orgyia pseu  
  
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W  
x[102])(L)x(S)(P)xx(L)(C)(K)xx(L)xxx(I)x(125f)  
MGGLVAFETTFENDMKARKIOSAREHYEDLATDROCRPDDEVKNDLSWKMYPKRPAPTLL  
DSKIKMSDYVHGRIAGIAIVFCIMGJGTMRINEAROLSDVLDANLIKKGKIRS  
DTIGLRKRNRKNTLNNKTPELARIVARNPTVQISKNSTSPKDFRLLDGVEMERSRSMIHLYLSNLYNSGVPLQ

S78187 ck: 3067 len: 197 | ymf39 protein - Reclinomonas americana (  
  
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W  
x[97])(I)x(S)(L)xx(F)(T)(K)xx(V)xxx(I)x(84)  
MUSFIONKHVTLILFLVALIHLDAHDFIIVHDETIVLDCFIILFFLYVALKDXYTASENDRA  
GIRILEGAISLQDKEEANS

T11851 ck: 3584 len: 324 | ruvb Protein - Thermus aquaticus thermop  
  
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W  
x[149])(L)x(T)(A)xx(L)(S)(R)xx(I)xxx(L)[159]  
VBDLALRPKTDDEYIGOERLQKQLREYEAAAKAREPEHLULLFGPGLGKTTLAHVIAHELGS  
MRDARLLGVRTTEEALEIGRRSRGGTYARRVDFNQAGEEYIIRERALE  
ALMALGIDELGLEKREDILEVILLIRFGAGPVGLATIALATLSEDPGLLEEYHEPYLLROGLKTRTPRGVATELAYRHLYPRPV

JC5705 ck: 3534 len: 125 | alpha-amylase inhibitor 0.19 - synthetic  
  
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W  
x[92])(A)x(S)(I)xx(V)(C)(R)xx(I)xxx(A)x(17)  
MSGPMWCYPCGAFQVPALPACRCRLLRLRQCNGSQVPEAVLDDCCOQLAHISEMRCGALYSM

Databases searched:  
MBRF, Release 62.0, Released on 30sep1999, Formatted on 26oct1999

Total finds: 561  
Total lengths: 47,169,319  
Total sequences: 142,080  
CPU time: 05:25.66

1: A SEQUENCE 1.0  
1: COGAT - cytochrome c2 - Rhodocyclus tenuis

C:Species: Rhodocyclus tenuis  
C>Date: 24-Apr-1984 #sequence\_revision 24-Apr-1984 #text\_change 15-Jan-1999  
C:Accession: A00090  
R:Ambler, R.P.; Meyer, T.E.; Kamen, M.D.  
A:Title: Anomalies in amino acid sequences of small cytochromes c and cytochromes c' from two species of purple photosynthetic bacteria.  
A:Reference number: A93207; MUID:79199668  
A:Accession: A00090  
A:Molecule type: protein  
A:Residues: 1-92 <AMB>  
C:Comment: This sequence is more closely related to the sequences of cytochrome c551 from Pseudomonas and Azotobacter than to the sequences of cytochrome c2 from other species of Rhodospirillum.  
C:Superfamily: cytochrome c6; cytochrome c6 homolog  
C:Keywords: chromoprotein; electron transfer; heme; iron; photosynthesis  
F:1-83/Domain: cytochrome c6 homology <CYC>  
F:12,15/Binding site: heme (cys) (covalent) #status predicted  
F:16,66/Binding site: heme iron (His, Met) (axial ligands) #status predicted  
CCOF2P Length: 92 February 11, 2000 15:51 Type: P Check: 2620 ..

1 ADESLAOTK GCLACHNEK KVGPAVGV AKRYAGAGA EARLVAKYA  
51 GGQGVAKOL GAELPMANN VKTEATRLV KWLISLKQID YK

!!A\_SEQUENCE 1.0  
P1:RDHCB - carbonyl reductase (NADPH) (EC 1.1.1.184) - human  
N:Alternate names: aldehyde reductase I; NADPH-dependent carbonyl reductase;  
Postaglandin 9-ketoreductase; xenobiotic ketone reductase  
C:Species: Homo sapiens (man)  
C:Accession: M61271; 831912; S08081; A47326  
R:Forrest, G.; Akman, S.; D'Show, J.; Rivera, H.; Kaplan, W.D.  
Mol. Pharmacol. 40: 502-507, 1991  
A:Title: Genomic sequence and expression of a cloned human carbonyl reductase gene with daunorubicin reductase activity.  
A:Reference number: A61271; MUID:92017676  
A:Accession: A61271  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-277 <FOR>  
A:Cross-references: GB:M62420; NID:q179977; PIDN:AA17881.1; PID:q179978  
R:Wernuth, B.; Bohren, K.M.; Heinemann, G.; von Wartburg, J.P.; Gabbay, K.H.  
J. Biol. Chem. 263: 16185-16188, 1988  
A:Title: Human carbonyl reductase. Nucleotide sequence analysis of a cDNA and amino acid sequence of the encoded protein.  
A:Reference number: A31912; MUID:89034082  
A:Accession: A31912  
A:Molecule type: mRNA  
A:Residues: 1-277 <WER>  
A:Cross-references: GB:J04056; NID:q181036; PIDN:AA52070.1; PID:q181037  
A:Note: part of this sequence was confirmed by protein sequencing  
R:Forrest, G.L.; Akman, S.; Krutzyk, S.; Paxton, R.J.; Sparkes, R.S.; Doroshov, J.; Felsted, G.L.; Glover, C.J.; Mohandas, T.; Bachur, N.R.  
Biochim. Biophys. Acta 1048: 149-155, 1990  
A:Title: Induction of a human carbonyl reductase gene located on chromosome 21.  
A:Reference number: S09013; MUID:90212644  
A:Accession: S09013  
A:Molecule type: mRNA  
A:Residues: 1-277 <FO>  
A:Cross-references: GB:J04056; EMBL:X51818; NID:q181036; PIDN:AA52070.1; PID:q181037  
R:Krook, M.; Ghosh, D.; Streemborg, R.; Carlgust, M.; Joernvall, H.  
Proc. Natl. Acad. Sci. U.S.A. 90: 502-506, 1993  
A:Title: Carboxyethyllysine in a protein: native carbonyl reductase/NADP-dependent prostaglandin dehydrogenase.  
A:Reference number: A47326; MUID:93133816  
A:Accession: A47326  
A:Molecule type: protein  
A:Residues: 221-242 <KRO>  
A:Note: Identification of carboxyethyllysine modification  
C:Comment: This cytosolic aldo-ketoreductase and quinone reductase reduces a

variety of carbonyl compounds to the corresponding alcohols.  
C:Genetics:  
A:Gene: GDB:CBR  
A:Cross-references: GDB:126610; OMIM:114830  
A:Map position: 21q22.12-21q22.12  
A:Introns: 97/1; 133/1  
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology  
C:Keywords: blocked amino end; monomer; NADP; oxidoreductase  
F:6-185/Domain: short-chain alcohol dehydrogenase homology <SADH>  
F:239/Modified site: N6-1-carboxyethyllysine (Lys) (partial) #status experimental

RDHCB Length: 277 February 11, 2000 15:51 Type: P Check: 7216 ..

1 MSSGHVALV TGSNGIGLA IVRLCRLES GDVVLIRADV TRGAAVOQL  
51 QAEGLSPRF QLDIDLOSI RALRDLRKE YGGLDVLYNN AGIAFKVADP  
101 TPFIOAEVT MKTNFGRD VCTELPLIK PGRVNVASS IMSVRLKSC  
151 SPEIQOKFRS ERTTEELWG LMKFVEDTK KGVHKEGWP SSAYGVTRIG  
201 VTVLSRIHAR KLSQRKDK ILNACCPCW VRTDMAGPRA TKSDEGAET  
251 PVYALPPD AEGPHQGEVS EKREVEOW

!!A\_SEQUENCE 1.0  
P1:RDALAE - acetoacetyl-CoA reductase (EC 1.1.1.36) - Alcaligenes eutrophus  
C:Species: Alcaligenes eutrophus  
C>Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 11-Jun-1999  
C:Accession: B34340  
R:Peoples, O.P.; Sinskey, A.J.  
J. Biol. Chem. 264: 15293-15297, 1989  
A:Title: Poly-beta-hydroxybutyrate biosynthesis in Alcaligenes eutrophus H16. Characterization of the genes encoding beta-ketothiolase and acetoacetyl-CoA reductase.  
A:Reference number: A34340; MUID:89359356  
A:Accession: B34340  
A:Molecule type: DNA  
A:Residues: 1-246 <PEO>  
A:Cross-references: EMBL:J04987; NID:q141953; PIDN:AA21973.1; PID:q141955  
A:Experimental source: strain H16  
C:Comment: This enzyme catalyzes the reduction of acetoacetyl-CoA to hydroxyacyl-CoA in the presence of NADPH, an essential step in the biosynthesis of poly-beta-hydroxybutyrate.  
C:Genetics:  
A:Gene: pubB  
A:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology  
C:Keywords: NAD; oxidoreductase; poly-beta-hydroxybutyrate biosynthesis  
F:4-184/Domain: short-chain alcohol dehydrogenase homology <SADH>  
RDALAE Length: 246 February 11, 2000 15:51 Type: P Check: 2829 ..

1 MQRIRAVYG GNGGIGTALC ORLAKDFRY VAGCCPSNPR REKLEODKA  
51 LGDFLASEG NWADMDSTKT AFDKXSEVG EVDVLIINAG ITRDVFRRM  
101 TRADMVAVID TWLISLENT KYVIDGADR GMGRIVNIS VNGCKGQFG  
151 TNSYAKAGL HGTMALAOE VARKGYVNT VSPGIADIM VKAIRDVIDL  
201 KIYATIPYKR LGPEELIASI CAWLSESESG PSTGADFEIN GGLHMG

!!A\_SEQUENCE 1.0  
P1:S05397 - granatidin polyketide ketoreductase (EC 1.1.1.-) graIII - Streptomyces violaceoniger  
N:Alternate names: 3-hydroxyacyl-CoA dehydrogenase homolog 1  
C:Species: Streptomyces violaceoniger  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: S05397  
R:Sherman, D.H.; Malpartida, F.; Bibb, M.J.; Kieser, H.M.; Bibb, M.J.; Hopwood, D.A.  
EMBO J. 8: 2717-2725, 1989

A>Title: Structure and deduced function of the granaticin-producing polyketide synthase gene cluster of *Streptomyces violaceoruber* Tue22.  
 A:Reference: Nucleic Acids Res. 27:5053-5059, 1999. PMID:9060034  
 A:Accession: U05393; MUID:9060034  
 A:Residue type: DNA  
 A:Molecule type: DNA  
 A:Residues: 1-372 <SMH>  
 A:Cross-references: EMBL:X16144; NID:947976; PIDN:CAA34263.1; PID:947978  
 C:Genetics:  
 C:Gene: grt11  
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology  
 C:Keywords: NAD; oxidoreductase  
 F:18-199/Domain: short-chain alcohol dehydrogenase homology <SADH>  
 505397 Length: 272 February 11, 2000 15:51 Type: P Check: 3104

1 MTATATATA TPTAAKFA LVTGATSGIG LAIARLAL GATFLCAGD  
 51 EERLQATKE LRGEFVDG TVCDVADPAQ IRAYVAAVQ RGTVDILVN  
 101 NAGRGGGAT AEIADLMID VITNTLSVF LMTKEVLNAG GMLAKRGRI  
 151 INIATGGKQ GVHAYPYSA SRKGVGLTK ALGELARTG ITVNAVCPGF  
 201 VETPAERYR EHYAGIQVS EERTEDRTN RVLGRVET REYAAVEXL  
 251 VADDAAVTA QALNVCGLG NY

11AA\_SEQUENCE 1.0  
 P1:S25079 - monensin polyketide ketoreductase (EC 1.1.1.-) - *Streptomyces cinamonensis*  
 C:Species: *Streptomyces cinamonensis*  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: S25079; S18170  
 R:Arrowsmith, T.J.; Malpartida, F.; Sherman, D.H.; Birch, A.; Hopwood, D.A.; Robinson, J.A.  
 Mol. Genet. 234, 254-264, 1992  
 A>Title: Characterisation of acti-homologous DNA encoding polyketide synthase genes from the monensin producer *Streptomyces cinamonensis*.  
 A:Reference number: S25076; MUID:92374994  
 A:Accession: S25079  
 A:Molecule type: DNA  
 A:Residues: 1-261 <ARR>  
 A:Cross-references: EMBL:Z15111; NID:946799; PIDN:CAA77599.1; PID:946803  
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology  
 C:Keywords: NAD; oxidoreductase  
 F:7-188/Domain: short-chain alcohol dehydrogenase homology <SADH>  
 S25079 Length: 261 February 11, 2000 15:51 Type: P Check: 4473

1 MGUSTSRVAL VTGATSGIGL ATARLLAAG HLVEFGARTE SVYATVYKAL  
 51 RNDGLEAEQ VLDVVDGASV TAFVQNAVDR YGRIDVLVNN AGRSGGCVTA  
 101 DLTLDELMDV IDTNLSVSR MTRAVLTGTG MTRERGRIT INASTGGKQG  
 151 VILGAPYSAS KHGVGFTKA LGNELAPGTI TVNAVCPGV ETMAQRVRO  
 201 GYAAVDTTE EAILTKFOAK IPIGRSTPE EVAGLIGTILA SDTAASITSO  
 251 ALNVCGLGN F

11AA\_SEQUENCE 1.0  
 P1:B42147 - 3-oxoacyl-[acyl-carrier-protein] reductase (EC 1.1.1.100) - *Escherichia coli*  
 N:Alternate names: 3-ketoacyl-ACP reductase  
 C:Species: *Escherichia coli*  
 C>Date: 30-Sep-1993 #sequence\_revision 31-Oct-1997 #text\_change 11-Jun-1999  
 C:Accession: B64853; B42147; C41856  
 R:Blattner, F.R.; Plunkett, G.; Bloch, C.A.; Berna, N.T.; Burland, V.; Riley, M.; Goland, P.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Laskin, J.; H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A>Title: The complete genome sequence of *Escherichia coli* K-12.

A:Reference number: A64720; MUID:97426617  
 A:Accession: B64853  
 A>Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-244 <BLAT>  
 A:Cross-references: GB:AE000210; GB:U00096; NID:91787332; PIDN:AACT4177.1; PID:91787335; UMG:U00096  
 A:Experimental source: strain K-12, substrain MG1655  
 R:Rawlings, M.; Cronan Jr., J.E.  
 J. Biol. Chem. 267, 5751-5754, 1992  
 A>Title: The gene encoding *Escherichia coli* acyl carrier protein lies within a cluster of fatty acid biosynthetic genes.  
 A:Reference number: A42147; MUID:92210530  
 A:Accession: B42147  
 A:Molecule type: DNA  
 A:Residues: 1-29, 6', 31-244 <RAM>  
 A:Cross-references: GB:M64991; NID:9145879; PIDN:AAA23739.1; PID:9145881  
 R:Verwoert, I.I.; Verdree, E.C.; van der Linden, K.H.; Nijkamp, H.J.; Stultje, A.R.  
 J. Bacteriol. 174, 2851-2857, 1992  
 A>Title: Cloning, nucleotide sequence, and expression of the *Escherichia coli* *fabD* gene, encoding malonyl coenzyme A-acyl carrier protein transacylase.  
 A:Reference number: A41856; MUID:92234941  
 A:Accession: C41856  
 A:Molecule type: DNA  
 A:Residues: 1-42, 44-45 <VER>  
 A:Cross-references: GB:M67040; NID:9145885; PIDN:AAA23743.1; PID:9145888  
 A>Note: sequence inconsistent with the nucleotide translation (NCBI:97150)  
 C:Genetics:  
 A:Gene: *fabG*  
 A:Map position: 24 min  
 C:Function: fatty acid biosynthesis  
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology  
 C:Keywords: fatty acid biosynthesis; NAD; oxidoreductase  
 F:6-18/Domain: short-chain alcohol dehydrogenase homology <SADH>  
 F:6-36/Region: beta-alphabeta NAD nucleotide binding fold  
 F:151/Active site: Tyr #status Predicted  
 B42147 Length: 244 February 11, 2000 15:51 Type: P Check: 3492

1 MNEGRKALV TGASRGIRA IAEITLARGA KYIGRTSEN GAAIDYIG  
 51 ANKGMLNV TDPASIEVL EKIRAFGEV DILVNNAGT RDLNLRMD  
 101 EEMNDILFEN LSVPRLSKA VMRAAMKRRH GRITIGSVV GIVNGGGOAN  
 151 YAAKAGLIG PEKSLAREVA SRGITVYVA PGFIETDMTR ALSDDGRAGI  
 201 LAQVPAGRLG GAOELANVA FLASDEAVI TGETLHVNG MYWV

11AA\_SEQUENCE 1.0  
 P1:D69930 - probable 3-oxoacyl-[acyl-carrier-protein] reductase (EC 1.1.1.100) *Xydoc - Bacillus subtilis*  
 C:Species: *Bacillus subtilis*  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: D69930; S01270  
 R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azavedo, J.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borcherdt, S.; Boursier, L.; Brans, A.; Braun, M.; Brignelli, S.C.; Bron, S.; Brock, L.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.; Conerton, J.F.; Cummings, N.J.; Daniel, R.A.; Demitro, F.; Devine, K.M.; Duestcher, A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandl, G.; Guiseppl, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Itaya, M.; Jones, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Nock, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Pottebell, D.; Potwolk, S.; Prescott, A.M.; Pressac, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Ray, M.; Reynolds, T.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, S.; Scallion, E.

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serro, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Taccioni, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenberg, M.; Vanherle, F.; Vassaret, A.; Viari, A.; Wamburt, R.; Wedler, E.; Wedler, H.; Weisenger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A:Reference number: A69580; MUID:98044033

A:Accession: D69930

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1238 <KUN>

A:Cross-references: GB:299114; GB:AL009126; NID:92634230; PIDN:CB13743.1; PID:EL15323; PID:92634244

A:Experimental source: strain 168

R:Cartigan, C.M.; Haasma, J.A.; Smith, M.T.; Wake, R.G.

Nucleic Acids Res. 15, 8501-8509, 1987

A:Title: Sequence features of the replication terminus of the *Bacillus subtilis* chromosome.

A:Reference number: S01270; MUID:88040469

A:Accession: S01270

A:Molecule type: DNA

A:Residues: 62-338 <CAR>

A:Cross-references: EMBL:X06168; NID:940205; PIDN:CAA29533.1; PID:9809662

C:Genetics:

A:Gene: yoxD

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

C:Keywords: NAD; oxidoreductase

F:7186/Domain: short-chain alcohol dehydrogenase homology <SADH>

D69930 Length: 238 February 11, 2000 15:51 Type: P Check: 2171

1 MOSLOKHTAL ITGGGRGIGR ATALALAKES VNIGLIGRTS ANEKVAEEV

51 KALGVAAFA ADVKADADQV NQAVQYKEG LGDIDILINN AGSKFGGFL

101 DLSADEMENT IOVNLGMYTH VTRAVLPEMI ERKADILINI SSTAQRGA

151 VTSNYSASKE AVLGITELSM QEVKKNHIV SALTSTYAS DMSIELNLTD

201 GNPEKWOPE DIAETVMAQL KLDPRIFIKT AGLNSTNP

11AA\_SEQUENCE 1.0

P1:DECEP - N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38) - *Escherichia coli*

C:Species: *Escherichia coli*

C:Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 11-Jun-1999

C:Accession: J03032; A42377; A65203; A50776

R:Parrot, C.; Boyen, A.; Cohen, G.N.; Glansdorff, N.

Gene 68, 275-283, 1988

A:Title: Nucleotide sequence of *Escherichia coli* argB and argC genes: comparison of N-acetylglutamate kinase and N-acetylglutamate-semialdehyde dehydrogenase with homologous and analogous enzymes.

A:Reference number: J03031; MUID:89121510

A:Accession: J03032

A:Molecule type: DNA

A:Residues: 1-334 <PAR>

A:Cross-references: GB:M21446; NID:9145332; PIDN:AA23477.1; PID:9145333

R:Meinert, T.; Schmitt, E.; Mechlam, Y.; Blanquet, S.

J. Bacteriol. 174, 2323-2331, 1992

A:Title: Structural and biochemical characterization of the *Escherichia coli* argB gene product.

A:Reference number: A42377; MUID:92202162

A:Accession: A42377

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-19 <MEI>

A:Cross-references: GB:X55417

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of *Escherichia coli* K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: A65203

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-334 <BLAT>

A:Cross-references: GB:AE000470; GB:U00096; NID:92367332; PIDN:AACT6940.1; PID:91790396; UMG:9398

A:Experimental source: strain K-12, substrain MG1655

C:Comment: In arginine biosynthesis glutamate is first converted to N-acetylglutamate, which leads to N-acetylglutamylphosphate and then to acetylglutamate kinase and N-acetyl-gamma-glutamyl-phosphate reductase.

C:Genetics:

A:Gene: argC

A:Map position: 90 min

C:Superfamily: N-acetyl-gamma-glutamyl-phosphate reductase

C:Keywords: arginine biosynthesis; oxidoreductase

F:154/Active site: Cys #status predicted

DECEP Length: 334 February 11, 2000 15:51 Type: P Check: 6342

1 MENTLVGAS GYAGAEVTV VNRHPHMT ALTVASQSD AGKLSDLHP

51 QKGIYDPL QPMDSIEFS PGVDVFLAT AHEVSHDLP OLEAGCVF

101 DLSGARVND ATFEKXYGF THOPELLEQ AAVGLAEMG NTKKANLIA

151 VPCCTPAAQ LALPPLDAD LLDINQWPI NATGVSAG RAAISNSPC

201 EYSLQYGF THROPELAT HIGADYIFP HLGFPFGIL EITTCRLSG

251 VTQAYAVL QQAVAHKPLV RLYDKGVAL KNYGLPFCF IGFVNGEHL

301 IIVATEDNLL KGAQAQAVOC ANIRFGYAEI QSLI

11AA\_SEQUENCE 1.0

P1:DECC - aspartate carboxyltransferase (EC 2.1.3.2) catalytic chain - *Escherichia coli* (strain K-12)

N:Alternate names: aspartate transcarboxylase catalytic chain; aspartyl carbamoyltransferase catalytic chain; carbamylaspartotransferase catalytic chain

C:Species: *Escherichia coli*

C:Date: 30-Nov-1980 #sequence\_revision 10-Oct-1997 #text\_change 11-Jun-1999

C:Accession: H65236; A00561; A21121; B36599; A21120; S56471

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of *Escherichia coli* K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: H65236

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-311 <BLAT>

A:Cross-references: GB:AE000495; GB:U00096; NID:92367361; PIDN:AACT7202.1; PID:92367364; UMG:94245

A:Experimental source: strain K-12, substrain MG1655

R:Schachman, H.K.; Paua, C.D.; Navre, M.; Karels, M.J.; Wu, L.; Yang, Y.R.

Proc. Natl. Acad. Sci. U.S.A. 81, 115-119, 1984

A:Title: Location of amino acid alterations in mutants of aspartate transcarboxylase: structural aspects of interallelic complementation.

A:Reference number: A93985; MUID:84119419

A:Accession: A00561

A:Molecule type: DNA

A:Residues: 1-149, 'E', 151-311 <SCH>  
 A:Cross-references: GB:K01472; NID:9147463; PIDN:AAA24476.1; PID:9147464  
 R:Hoover, T.A.; Roof, W.D.; Foltermann, K.F.; O'Donovan, G.A.; Bencini, D.A.; Wild, J.R.  
 Proc. Natl. Acad. Sci. U.S.A. 80, 2462-2466, 1983  
 A:Title: Nucleotide sequence of the structural gene (pyrB) that encodes the catalytic polypeptide of aspartate transcarbamoylase of *Escherichia coli*.  
 A:Reference number: A21121; MUID:83195078  
 A:Accession: A21121  
 A:Molecule type: DNA  
 A:Residues: 1-60, 'Q', 62-165, 'T', 167-220, 'V', 222-311 <HOOS>  
 A:Cross-references: GB:V00323  
 R:Donahue, J.P.; Turnbough Jr., C.L.  
 J. Biol. Chem. 265, 19091-19099, 1990  
 A:Title: Characterization of transcriptional initiation from promoters P-1 and P-2 of the pyrB operon of *Escherichia coli* K12.  
 A:Reference number: A36599; MUID:91035438  
 A:Accession: B36599  
 A:Molecule type: DNA  
 A:Residues: 1-18 <DON>  
 A:Cross-references: GB:M60508; NID:9147469; PIDN:AAA24481.1; PID:9147471  
 R:Konysberg, W.H.; Henderson, L.  
 Proc. Natl. Acad. Sci. U.S.A. 80, 2467-2471, 1983  
 A:Title: Amino acid sequence of the catalytic subunit of aspartate transcarbamoylase from *Escherichia coli*.  
 A:Reference number: A21120; MUID:83195079  
 A:Accession: A21120  
 A:Molecule type: protein  
 A:Residues: 2-60, 'Q', 62-86, 'Q', 88-90, 'N', 92-129, 'N', 131-220, 'V', 222-256, 'D', 258-259, 'M', 261, 'A', 263-311 <KON>  
 R:Ke, H.; Honzato, R.B.; Lipscomb, W.N.  
 Proc. Natl. Acad. Sci. U.S.A. 81, 4037-4040, 1984  
 A:Title: Structure of unliganded aspartate carbamoyltransferase of *Escherichia coli* at 2.6-angstroms resolution.  
 A:Reference number: A93993; MUID:84248054  
 A:Contents: annotation; X-ray crystallography, 2.6 angstroms; quaternary structure  
 R:Burkett, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.  
 Nucleic Acids Res. 23, 2109-2119, 1995  
 A:Title: Analysis of the *Escherichia coli* genome VI: DNA sequence of the region from 92.8 through 100 minutes.  
 A:Reference number: S56314; MUID:95334362  
 A:Accession: S56471  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-195, 'R', 197-311 <BUR>  
 A:Cross-references: EMBL:U14003; NID:91263172; PIDN:AA97142.1; PID:9537087  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994  
 C:Comment: The active enzyme contains two trimers of catalytic chains and three dimers of regulatory chains; it catalyzes the rate-limiting step in the biosynthesis of pyrimidines.  
 C:Genetics:  
 A:Gene: pyrB  
 A:Map position: 97 min  
 C:Superfamily: ornithine carbamoyltransferase; aspartate/ornithine carbamoyltransferase homology  
 C:Keywords: heterododecamer; homohexamer; homotrimer; pyrimidine nucleotide biosynthesis; transerase  
 F:8-305/Domain: aspartate/ornithine carbamoyltransferase homology <ACT>  
 DECC Length: 311 February 11, 2000 15:51 Type: P Check: 4302

1 MAPLXOKHI ISINDSRD LNLVLTAAK LKANPOPEL HKVYASCF  
 51 EASTRRRLSF ETSMRRLGAS VVGFSDSANT SLGKKGELTA DTIVISTYV  
 101 DAIVMHPDE GAARLATEFS GNVPLVAGD GSNQHPQTQL LDLFTIETQ  
 151 GRDNIHIM VGDLYKGRIV HSLTQALAK DGNRFYFIAP DALAMPQYI  
 201 DMDEKGIAM SLHSIEEV AVVDILYMT VQKERLPSE YANVKAQYV

251 RASDLHNKA NKVYHPLPR VDEIATDVCK TPNAVFOOA GNGIFAQAL  
 301 LALVNSRDIV L  
 I:AA\_SEQUENCE 1.0  
 P1:OWBAC - aspartate carbamoyltransferase (EC 2.1.3.2) catalytic chain - *Salmonella typhimurium*  
 N:Alternate names: aspartate transcarbamylase catalytic chain; carbamylaspartotranskinase  
 C:Species: *Salmonella typhimurium*  
 C:Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 11-Jun-1999  
 A:Accession: S00049  
 R:Michaels, G.; Kelln, R.A.; Nargang, F.E.  
 Eur. J. Biochem. 166, 55-61, 1987  
 A:Title: Cloning, nucleotide sequence and expression of the pyrB operon of *Salmonella typhimurium* LT2.  
 A:Reference number: S00028; MUID:87246692  
 A:Accession: S00049  
 A:Molecule type: DNA  
 A:Residues: 1-311 <MIC>  
 A:Cross-references: GB:X05641; NID:947861; PIDN:CAA29129.1; PID:947863  
 C:Comment: The active enzyme contains two trimers of catalytic chains and three dimers of regulatory chains; it catalyzes the rate-limiting step in the biosynthesis of pyrimidines.  
 C:Genetics:  
 A:Gene: pyrB  
 A:Map position: 98 min  
 C:Superfamily: ornithine carbamoyltransferase; aspartate/ornithine carbamoyltransferase homology  
 C:Keywords: heterododecamer; homohexamer; homotrimer; pyrimidine nucleotide biosynthesis; transerase  
 F:8-305/Domain: aspartate/ornithine carbamoyltransferase homology <ACT>  
 OMBAC Length: 311 February 11, 2000 15:51 Type: P Check: 4073

1 MAPLXOKHI ISINDSRD LNLVLTAAK LKANPOPEL HKVYASCF  
 51 EASTRRRLSF ETSMRRLGAS VVGFSDSANT SLGKKGELTA DTIVISTYV  
 101 DAIVMHPDE GAARLATEFS GNVPLVAGD GSNQHPQTQL LDLFTIETQ  
 151 GRDNIHIM VGDLYKGRIV HPAKPTLAK ESGNRFYFIA PDALAMPQYI  
 201 LMDLDEKMA WSLGSIIEV MADVDILYMT RVQKERLPDS EYANVKAQYV  
 251 LRPLDNGARE NKVYHPLPR IDEITDVCK TPNAVFOOA GNGIFAQAL  
 301 LALVNSRDL L  
 I:AA\_SEQUENCE 1.0  
 P1:OWBAC - aspartate carbamoyltransferase (EC 2.1.3.2) catalytic chain - *Serratia marcescens*  
 N:Alternate names: aspartate transcarbamylase catalytic chain; carbamylaspartotranskinase  
 C:Species: *Serratia marcescens*  
 C:Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 11-Jun-1999  
 A:Accession: B34396  
 R:Beck, D.; Redzie, K.W.; Wild, J.R.  
 J. Biol. Chem. 264, 16629-16637, 1989  
 A:Title: Comparison of the aspartate transcarbamoylases from *Serratia marcescens* and *Escherichia coli*.  
 A:Reference number: A34396; MUID:89380286  
 A:Accession: B34396  
 A:Molecule type: DNA  
 A:Residues: 1-306 <BECK>  
 A:Cross-references: GB:J05033; NID:9398074; PIDN:AA26564.1; PID:9398075  
 C:Comment: The active enzyme contains two trimers of catalytic chains and three dimers of regulatory chains; it catalyzes the rate-limiting step in the biosynthesis of pyrimidines.  
 C:Genetics:  
 A:Gene: pyrB  
 C:Superfamily: ornithine carbamoyltransferase; aspartate/ornithine



carbamoyltransferase homology  
 C:Keywords: heterododecamer; homohexamer; homotrimer; pyrimidine nucleotide biosynthesis; transference  
 F:8-300/Domain: aspartate/orntithine carbamoyltransferase homology <ACT>  
 OMSAC Length: 306 February 11, 2000 15:51 Type: P Check: 6543 ..

1 MANPLYHKHI ISINDLRSD LEVLATPAAG LKANPOPELL KHKYIASCFE  
 51 EASTRRLRSE ETSMHRLGAS VYGFADGSGNT SLCKKGETIA DTISVISTYV  
 101 DAIVMRHPQE GAWMASEFSG NNPVLANAGDG NQHPOTOTLID LFTIETQGR  
 151 LSNLSIANVG DAKYGRVHS LQALAKREG NREYFIAPDA LAMPAYILKM  
 201 LEERKEEYS HGSEIEVPE LDILMTVQ KERLDPESEA NKKAQFVLAA  
 251 DIAGANLKV LHPRLRDEI AIDVKTIPA YTFQAGNGI FARSALALV  
 301 NADLAL

!!AA:SEQUENCE 1.0  
 P1:OWP - ornithine carbamoyltransferase (EC 2.1.3.3) - fission yeast  
 (Schizosaccharomyces pombe)  
 N:Alternate names: citrulline phosphorlyase; ornithine transcarbamylase  
 C:Species: Schizosaccharomyces pombe  
 C>Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 11-Jun-1999  
 C:Accession: S22390; S22175  
 Eur. J. Biochem. 205, 33-43, 1992  
 Rivan Hufel, C.; Dubois, E.; Messenguy, F.  
 A:Title: Cloning and sequencing of arg3 and arg1 genes of Schizosaccharomyces pombe on a 10-kb DNA fragment. Heterologous expression and mitochondrial targeting of their translation products.  
 A:Reference number: S22389; MUID:92209520  
 A:Accession: S22390  
 A:Molecule type: DNA  
 A:Residues: 1-327 <VAV>  
 A:Cross-references: EMBL:X63577; NID:94907; PIDN:CAA45133.1; PID:94908  
 C:Genetics: arg3  
 A:Gene: arg3  
 C:Map position: 1L  
 C:Superfamily: ornithine carbamoyltransferase; aspartate/orntithine carbamoyltransferase homology  
 C:Keyword: arginine biosynthesis; mitochondrion; transference  
 F:8-315/Domain: aspartate/orntithine carbamoyltransferase homology <ACT>  
 OWP Length: 327 February 11, 2000 15:51 Type: P Check: 9240 ..

1 MSRRKPRHL ISIRLSRGE IVKLIDRSF IKQAKQNFQ NRRSQMSG  
 51 SSONVAMIFS KRSTRVSV ESAVSLGSGN AMFLKRDIO LGVNESLYOT  
 101 SKVSSKWSG IVARVNTYSD VATLAKHASC PYINGLCOTF HPLOALADLL  
 151 TIKETEKSPD GLKVAWVGA NNVLHDMIA NAKVGIHTSV AKRPDVNRD  
 201 DILSIWNEA NENGSTFEIV NDKPVAVKNA DIVVTDTWIS MGOFAEKOR  
 251 LKQFTGQVY GEIMRLKPS CKFMHCLPRH PEVSDEVYV GENSLVFOEA  
 301 ENRKWTVAV LBALVNRGE ILPPASA

!!AA:SEQUENCE 1.0  
 P1:A69962 - phosphate butyryltransferase (EC 2.3.1.19) yqis - Bacillus subtilis  
 C:Species: Bacillus subtilis  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: A69962  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Bester, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boudster, L.; Brans, A.; Braun, M.; Bignelli, S.C.; Bron, S.; Brouillette, S.; Bruch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.; Comercon, I.F.; Cummings, N.T.; Daniel, R.A.; Denzot, F.; Devine, K.M.; Duesterhoft, A.; Ehrlich, S.D.; Emerson, P.T.; Ertlan, K.D.; Erlington, J.;

Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.; Guisepi, G.; Guy, B.J.; Hada, K.; Halech, J.; Harwood, C.R.; Hent, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kaashara, Y.; Kraier-Blanchard, M.; Klein, C.; Kodiyandi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapius, A.; Lardinois, S.  
 A:Authors: Lander, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Maude, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Mostl, D.; Nakai, S.; Noack, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogikawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porollik, S.; Prescott, A.M.; Prescan, E.; Pujo, P.; Fumelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sedate, Y.; Sato, T.; Scanlon, E.  
 A:Authors: Schleich, S.; Schroeter, R.; Scifone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serro, P.; Shin, B.S.; Solio, B.; Sorokin, A.; Taccoli, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Yamashita, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandendool, K.; Vanlier, F.; Vassariotti, A.; Viari, A.; Wambull, R.; Wedler, E.; Wedler, H.; Wetzenecker, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
 A:Reference number: A69580; MUID:98044033  
 A:Accession: A69962  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-299 <KUN>  
 A:Cross-references: GB:299116; GB:AL009126; NID:92634723; PIDN:CAB14340.1; PID:9185677; PID:92634843  
 A:Experimental source: strain 168  
 C:Genetics: yqis  
 A:Gene: yqis  
 C:Superfamily: phosphate acetyltransferase  
 C:Keywords: acyltransferase  
 A69962 Length: 299 February 11, 2000 15:51 Type: P Check: 158 ..

1 MKIKDIGRA SIHKNTIAY AHAEDEVIR AVKLAHEHS AFLLTGDSK  
 51 KINELTSMQ GHVEIYHAN TEPSKIAV RAVHKTADV LMKGVPTSV  
 101 LKAVLNROE GLRSAYLSH VAVFDIPDP RLMEFTDSAM NIAPSELEL  
 151 QILONAVHYA HAVGNMPPKA AALAAVEYV PMEATVNA AALQMKKQ  
 201 IKGCIYDGL ALDNAVSOIA AAKKISGDV AGNDILLV TIEAGNIIYK  
 251 SLIFAKASV AAVITGAKAP IALTSRADSA ENKLYSIALA ICASEEYTH

!!AA:SEQUENCE 1.0  
 P1:XNEBHC - histidinol-phosphate transaminase (EC 2.6.1.9) - Salmonella typhimurium  
 N:Alternate names: histidinol-phosphate aminotransferase  
 C:Species: Salmonella typhimurium  
 C>Date: 30-Sep-1990 #sequence\_revision 30-Sep-1990 #text\_change 28-May-1999  
 C:Accession: U00158  
 R:Carlomagno, M.S.; Chiarotli, L.; Alfano, P.; Napo, A.G.; Bruni, C.B.  
 J. Mol. Biol. 203, 585-606, 1988  
 A:Title: Structure and function of the Salmonella typhimurium and Escherichia coli K-12 histidine operons  
 A:Reference number: J00131; MUID:89094829  
 A:Accession: U00158  
 A:Molecule type: DNA  
 A:Residues: 1-359 <CAR>  
 A:Cross-references: GB:X13464; NID:947719; PIDN:CAA31824.1; PID:9477723  
 C:Comment: This enzyme catalyzes the conversion of imidazolephosphate to L-histidinol phosphate, the eighth step in histidine biosynthesis.  
 C:Genetics: hisC  
 A:Gene: hisC

A:Map position: 42 min  
C:Superfamily: histidinol-phosphate aminotransferase  
C:Keywords: aminotransferase; histidine biosynthesis

XNBBHC Length: 359 February 11, 2000 15:51 Type: P Check: 1849

1 MSTEINTSVA DLARENENML VPYOSARLG GNGDVLNAN EPTAVEFOL

51 TQOGLNRYPE COPKAVIENY AQYAGVKPEO VLVSRCADRG TELVIRFAC

101 PGKRALIYCP PTYGMYSVA ETICVERTV PALENQOLDI QGISNDLDT

151 KVFVVCSPNN PTGOLINPDD LRTLELRLRG KALVADEAV IEFQCAATLI

201 GMLVEYPHLV IRLTSLKAFK LAGLRGFTL ANEEVINLL KVIAPYPLST

251 PVADIAAQS CPOESNAMD RVAOTVDERQ YLVNALQOTA CVENFDESET

301 NYLIARFTAS SSVFNSLMDQ GIIRDONKO PSLSGCLRTIT VGTROENOPV

351 IDALRAEPV

!IAA\_SEQUENCE 1.0

P1:S29090 - dual specificity phosphoprotein phosphatase (EC 3.1.3.-) 1 - human

N:Alternate names: protein-tyrosine-phosphatase Cl100;  
protein-tyrosine-phosphatase, nonreceptor type 10

C:Species: Homo sapiens (man)  
C:Date: 25-Feb-1994 #sequence\_revision 02-May-1994 #text\_change 11-Jun-1999

C:Accession: S29090; A53052

R:Keyse, S.M.; Emslie, E.A.  
Nature 359, 644-647, 1992

A:Title: Oxidative stress and heat shock induce a human gene encoding a  
protein-tyrosine phosphatase.

A:Reference number: S29090; MID:93024952

A:Accession: S29090

A:Molecule type: mRNA

A:Residues: 1-367 <KEY>

A:Cross-references: EMBL:X68277; NID:g29980; PIDN:CAA6838.1; PID:g29981

R:Kwak, S.P.; Hakes, D.J.; Matteil, K.D.; Dixon, J.E.  
J. Biol. Chem. 269, 3596-3604, 1994

A:Title: Isolation and characterization of a human dual specificity  
protein-tyrosine phosphatase gene.

A:Reference number: A53052; MID:94148864

A:Accession: A53052

A:Molecule type: DNA

A:Residues: 1-367 <KWA>

A:Experimental source: leukocyte

A:Note: sequence extracted from NCBI backbone (NCBIN:143800, NCBIN:143802,  
NCBIN:143804, NCBIN:143806, NCBI:143807)

C:Genetics:

A:Gene: GDB:DUSP1; PTN10

A:Cross-references: GDB:136197; OMIM:600714

A:Map position: 5q34-5q34

C:Superfamily: dual specificity phosphoprotein phosphatase 1; VHL-type dual  
specificity phosphoprotein phosphatase homology

C:Keywords: heat shock; phosphoprotein; phosphoric monoester hydrolase;  
stress-induced protein

F:101-312/Domain: VHL-type dual specificity phosphoprotein phosphatase homology  
<VHL>  
F:228/Active site: Cys (phosphocysteine intermediate) #status predicted  
F:264/Binding site: substrate phosphate (Arg) #status predicted

S29090 Length: 367 February 11, 2000 15:51 Type: P Check: 5401

1 MMEVEGILDA GGLRALRGER AAOCLLDCR SFPAFAGHI AGSVNVRST

51 IYRRRAKAM GLEHIVPNAE LRGRLLAGAY HAVVILDEBS AALDCAKRDG

101 TLALAGALC REARAQVFF LKGYEAFSA SCEPLCSKOS TPMGLSLPLS

151 TSPVDSAESG CSSCSTPLYD OGGVEILPF LYGSAVYAS RKDMDALGI

201 TALINVSANC PNHFEHGYO KSIPEVDNKH ADLSSFNENA IDFIDISINA

251 GGRVYHCOA GISRSATICL AYLMRTNRK LDEAEFVKO RRSIISNFS

301 FMGOLLQFES OYLAPHCSE AGSPANAVLD RGTSTTVFN FVSPVHST

351 NSALSTYOSP ITTSPSC

!IAA\_SEQUENCE 1.0

P1:S24411 - dual specificity phosphoprotein phosphatase (EC 3.1.3.-) 1 - mouse

N:Alternate names: 3CH134 protein; protein-tyrosine-phosphatase erp,  
nonreceptor type 10

C:Species: Mus musculus (house mouse)  
C:Date: 19-Feb-1994 #sequence\_revision 02-May-1994 #text\_change 11-Jun-1999

C:Accession: A54681; S24411

R:Noguuchi, T.; Metz, R.; Chen, L.; Matteil, M.G.; Carrasco, D.; Bravo, R.  
Mol Cell Biol 13, 5195-5205, 1993

A:Title: Structure, mapping, and expression of erp, a growth factor-inducible  
gene encoding a nontransmembrane protein tyrosine phosphatase, and effect of  
erp on cell growth.

A:Reference number: A54681; MID:93360956

A:Accession: A54681

A:Molecule type: DNA

A:Residues: 1-367 <MOG>

A:Cross-references: GB:S64851; NID:g409976; PIDN:AAB27882.1; PID:g409977

R:Charles, C.H.; Adler, A.S.; Lau, L.F.  
Oncogene 7, 187-190, 1992

A:Title: CDNA sequence of a growth factor-inducible immediate early gene and  
characterization of its encoded protein.

A:Reference number: S24411; MID:92158357

A:Accession: S24411

A:Molecule type: mRNA

A:Residues: 1-367 <CHAS>

A:Cross-references: EMBL:X61940; NID:g49735; PIDN:CAA43944.1; PID:g49736

C:Genetics:

A:Gene: erp

A:Superfamily: 123/1; 172/1; 245/1

C:Superfamily: dual specificity phosphoprotein phosphatase 1; VHL-type dual  
specificity phosphoprotein phosphatase homology

C:Keywords: immediate-early protein; phosphoprotein; phosphoric monoester  
hydrolase

F:181-312/Domain: VHL-type dual specificity phosphoprotein phosphatase homology  
<VHL>  
F:238/Active site: Cys (phosphocysteine intermediate) #status predicted  
F:264/Binding site: substrate phosphate (Arg) #status predicted

S24411 Length: 367 February 11, 2000 15:51 Type: P Check: 5871

1 MMEVEGILDA GGLRALRREG AAOCLLDCR SFPAFAGHI AGSVNVRST

51 IYRRRAKAM GLEHIVPNAE LRGRLLAGAY HAVVILDEBS ASLDCARRDG

101 TLALAGALC REARSTQVFF LKGYEAFSA SCEPLCSKOS TPTGLSLPLS

151 TSPVDSAESG CSSCSTPLYD OGGVEILSF LYGSAVYAS RKDMDALGI

201 TALINVSANC PNHFEHGYO KSIPEVDNKH ADLSSFNENA IDFIDISKDA

251 GGRVYHCOA GISRSATICL AYLMRTNRK LDEAEFVKO RRSIISNFS

301 FMGOLLQFES OYLAPHCSE AGSPANAVLD RGTSTTVFN FVSPVHPT

351 NSALNTIKSP ITTSPSC

!IAA\_SEQUENCE 1.0

P1:PRMSCL - cytotoxic T-lymphocyte proteinase (EC 3.4.21.-) 1 precursor - mouse

N:Alternate names: CTLA-1 protein; cytotoxic cell proteinase 1 (CPT1);  
cytotoxic T-cell-specific proteinase 1; cytotoxic T-cell-specific proteinase  
granzyme B

C:Species: Mus musculus (house mouse)  
C:Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 18-Jun-1999

C:Accession: A94288; #sequence\_revision 04-Dec-1986 #text\_change 18-Jun-1999

R:Loeb, C.G.; Finlay, B.B.; Paranchych, W.; Paetkau, V.H.; Bleackley, R.C.  
Science 232, 858-861, 1986

```

A:Title: Novel serine proteases encoded by two cytotoxic T lymphocyte-specific
genes.
A:Reference number: A94288; MID:86208120
A:Accession: A94288
A:Molecule type: mRNA
A:Residues: 1-247 <LOB>
A:Cross-references: GB:M12302; NID:9192452; PIDN:AAA37383.1; PID:9309154
R:Brunet, J.F.; Dosseto, M.; Denizot, F.; Mattei, M.G.; Clark, W.R.; Haq,
T.M.; Ferrer, P.; Nabholz, M.; Schmitt-Vernhulst, A.M.; Luciani, M.F.;
Gustein, P. 268-271, 1986
A:Title: The inducible cytotoxic T-lymphocyte-associated gene transcript CTLA-1
sequence and gene localization to mouse chromosome 14.
A:Reference number: A93882; MID:86284960
A:Contents: chromosome mapping
A:Accession: A93882
A:Molecule type: mRNA
A:Residues: 1-247 <BRU>
A:Cross-references: EMBL:X04072; NID:950586; PIDN:CAA27715.1; PID:950587
R:Lobe, C.G.; Upton, C.; Dugan, B.; Letellier, M.; Bell, J.;
McFadden, G.; Blackley, R.C.
Biochemistry 27, 6941-6946, 1988
A:Title: Organization of two genes encoding cytotoxic T lymphocyte-specific
serine proteases CCP1 and CCP2.
A:Reference number: A90536; MID:89062424
A:Accession: A28952
A:Molecule type: DNA
A:Residues: 1-247 <LO2>
A:Cross-references: GB:M22526; NID:9226432; PIDN:AAB61756.1; PID:9201027
R:Masson, D.; Tschopp, J.
Cell 49, 679-685, 1987
A:Title: A family of serine esterases in lytic granules of cytolytic T
lymphocytes
A:Reference number: A90894; MID:87215932
A:Accession: B26944
A:Molecule type: protein
A:Residues: 21-40 <NAS>
R:Ko, M.S.; Wang, X.; Horton, J.H.; Hagen, M.D.; Takahashi, N.; Maeraki, Y.;
Nadeau, J.H.
Mamm. Genome 5, 349-355, 1994
A:Title: Genetic mapping of 40 cDNA clones on the mouse genome by PCR.
A:Reference number: I48934; MID:94319082
A:Accession: I48937
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 227-247 <RES>
A:Cross-references: EMBL:U05707; NID:9497035; PIDN:AAB60470.1; PID:9497036
C:Comment: This enzyme is probably necessary for target cell lysis in
cell-mediated immune responses.
C:Genetics:
A:Gene: CTLA-1
A:Map position: 14
A:Introns: 19/1; 68/2; 113/3; 200/3
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase; T-cell
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-20/Domain: propeptide #status predicted <APT>
F:21-24/Product: cytotoxic T lymphocyte proteinase 1 #status predicted <MPT>
F:21-24/Domain: trypsin homology <TRI>
F:49-65,142-209,173-188/Disulfide bonds: #status predicted
F:64,108,203/Active site: His, Asp, Ser #status predicted
PRMSL Length: 247 February 11, 2000 15:51 Type: P Check: 92 ..

1 MKILLILLTL SLASRTKAGE IIGGHEVRPH SRPYMALSLI KDQOEALICG
51 GFLIREDFVL TAAHCEGSI I NVTLGAHNIK EOERTQOIVP MVKCIPIHBDY
101 NPKTFSDIM LRLKSKAKR TRAVRPLNP RRVNVKPGD VCVYAGWGRM
151 APMGYSTNL QEVELFVQKD RECESYFKNR YNKTNQICAG DKPTRASFR
201 GDSGPIVCK KVAAGIVSYG YKDGSPPRAF TKVSSFLSWI KTKMSS

!!AA_SEQUENCE 1.0
P1:D5312 - cytidine deaminase (EC 3.5.4.5) - Mycoplasma pirum (strain BER)
(SGCC)
C:Species: Mycoplasma pirum
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: D53312
R:Tham, T.N.; Ferris, S.; Kovacic, R.; Montagnier, L.; Blanchard, A.
J. Bacteriol. 175, 5281-5285, 1993
A:Title: Identification of Mycoplasma pirum genes involved in the salvage
pathways for nucleosides.
A:Reference number: A53312; MID:93352438
A:Accession: D53312
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-133 <THA>
A:Cross-references: GB:L13289; NID:9401781; PIDN:AAA25433.1; PID:9401785
C:Genetics:
A:Gene: cdc
A:Genetic code: GCC3
A:Superfamily: cytidine deaminase
C:Keywords: hydrolase
D53312 Length: 133 February 11, 2000 15:51 Type: P Check: 6718 ..

1 MKKDIYQK INELISNAVY PYSNRFVSC LITDGFAG VIENSAYSP
51 TICAERSAVS SMITSGKQI FRVYILTDTI VKDIGPCGV CROVISEFAK
101 PETPIITNL KGEKFYTLLE QLLPFAFNKD ALK

!!AA_SEQUENCE 1.0

```

P1:JE0022 - cytidine/deoxycytidine deaminase (EC 3.5.4.-) cdd - *Bacillus subtilis*  
 N:Alternate names: cytidine 12'-deoxycytidine aminohydrolase; P43 protein  
 C:Species: *Bacillus subtilis*  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: JE0022; I39955; F69587; S04425  
 R:Song, B.H.; Neubard, J. 1989  
 Mol.Gen. Genet. 216: 462-468, 1989  
 A:Title: Chromosomal location, cloning and nucleotide sequence of the *Bacillus subtilis* cdd gene encoding cytidine/deoxycytidine deaminase.  
 A:Reference number: JE0022; M01D:89313687  
 A:Accession: JE0022  
 A:Molecule type: DNA  
 A:Residues: 1-136 <SON>  
 A:Cross-references: GB:018532; MID:9606743; PIDN:AAB59993.1; PID:9606744  
 R:Wang, P. 1984  
 J. Biol. Chem. 259, 8619-8625, 1984  
 A:Title: Overlapping promoters transcribed by *Bacillus subtilis* sigma-55 and sigma-37 RNA polymerase holoenzymes during growth and stationary phases.  
 A:Reference number: I39955; M01D:84239852  
 A:Accession: I39955  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-10 <RES>  
 A:Cross-references: GB:R02174; MID:9143349; PIDN:AAB05347.1; PID:9551724  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bartoe, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Connerion, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devigne, K.M.; Dubrethoff, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Farey, C.; Ferrari, E. 1997  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D., Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallenot, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandt, G.; Guseppl, G.; Guy, B.J.; Hage, K.; Halech, U.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hultio, M.F.; Itaya, M.; Jones, L.; Jorits, B.; Karamata, D.; Kasahara, Y.; Klier-Bianchard, M.; Klein, C.; Kodayashi, Y.; Koether, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.  
 A:Authors: Labber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Maestri, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moesti, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogizawa, A.; Ouega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelie, D.; Porwolk, S.; Prescott, A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.  
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serro, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Tepstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandebol, M.; Vannier, P.; Vassatotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Wetzinger, T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*.  
 A:Reference number: A69580; M01D:98044033  
 A:Accession: F69597  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-136 <KUN>  
 A:Cross-references: GB:299116; GB:299117; GB:AL009126; MID:92634966; PIDN:CAH4447.1; PID:el183760; PID:92634976; MID:92634723; PID:el185796; PID:92634962  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: cdd  
 A:Map position: 225 (degrees)  
 C:Superfamily: cytidine deaminase  
 C:Keywords: hydrolase  
 JE0022 Length: 136 February 11, 2000 15:51 Type: P Check: 9048 ..

1 MNR0ELINEA LKADNAYAP YSKFOVGAAL LTKDKYRG CNENAYSM  
 51 CNCAERTALF KAVSEGTPE QMLAAVADRP GPVSPGACR OVISELCTKD  
 101 VIVVLTINQG QIKEMTYEEL LPGAESSEDL HDEKRL  
 !!AA-SEQUENCE 1.0  
 P1:F69500 - riboflavin-specific deaminase (ribd) homolog - *Archaeoglobus fulgidus*  
 C:Species: *Archaeoglobus fulgidus*  
 C:Date: 10-Sep-1998 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: F69500  
 R:Kleink, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Keriavage, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenney, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L. 1997  
 Nature 390, 364-370, 1997  
 A:Authors: Ullrich, T.; Cotton, M.D.; Spriggs, T.; Arlath, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujita, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus*.  
 A:Reference number: A69250; M01D:98049343  
 A:Accession: F69500  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-219 <KLE>  
 A:Cross-references: GB:AE000964; GB:AE000782; MID:92669287; PIDN:AAB89247.1; PID:92648530; TIGR:AF2007  
 C:Superfamily: riboflavin-specific deaminase  
 F69500 Length: 219 February 11, 2000 15:51 Type: P Check: 4460 ..  
 1 MRPYFVAVVA ASLDGRKISDE SKRQRISE EDLKVDRLR AESDAIMVGI  
 51 GTVLADPRL TVKSALREK ROKDKEPNP LRVVDSRCR VPLTARLND  
 101 EARLVAISR IADPEKYREV KVAEVAVFG EERYELALL EFLHRGVRR  
 151 LMVEGGGLTI SSLISONLVD EIRIYGPIT IGRHDSPTVC DGBSEFLKCR  
 201 IEKIERIGEG FAVTARENR  
 !!AA-SEQUENCE 1.0  
 P1:I38612 - H+-transporting ATP synthase (EC 3.6.1.34) lipid-binding protein P3 precursor, mitochondrial - human  
 N:Alternate names: H+-transporting ATP synthase chain 9.3; H+-transporting ATP synthase chain c form P3  
 C:Species: *Homo sapiens* (man)  
 C:Date: 26-Aug-1999 #sequence\_revision 26-Aug-1999 #text\_change 26-Aug-1999  
 C:Accession: I38612  
 R:Ryan, W.L.; Ierner, T.J.; Haines, J.L.; Gusella, J.F.  
 Genomics 24, 375-377, 1994  
 A:Title: Sequence analysis and mapping of a novel human mitochondrial ATP synthase subunit 9 cDNA (ATP5G3).  
 A:Reference number: A55687; M01D:95213032  
 A:Accession: I38612  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-142 <YAN>  
 A:Cross-references: EMBL:U09813; MID:91008454; PIDN:AAA78807.1; PID:9511450  
 C:Genetics:  
 A:Gene: ATP5G3  
 A:Cross-references: GDB:375306  
 A:Map position: 2pter-2qter  
 A:Genome: nuclear  
 C:Superfamily: H+-transporting ATP synthase lipid-binding protein  
 C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrial; oxidative phosphorylation; transmembrane protein

F:1-67/Domain: transit peptide (mitochondrion) #status predicted <TRP>  
F:68-142/Product: H+-transporting ATP synthase lipid-binding protein #status  
Predicted <MAT>  
F:75-103/Domain: transmembrane #status predicted <TM1>  
F:113-142/Domain: transmembrane #status predicted <TM2>  
F:125/Active site: Glu #status predicted

I38612 Length: 142 February 11, 2000 15:51 Type: P Check: 2178 ..

1 MFACACIACI PSLIRAGSRV AYRISASVLR SREPSRICE GSTYENQAN  
51 GVSOLIERE OTSAISRDI TAAFTGAGA ATYGVAGSQA GIGVFQSLI  
101 IGVARNPSLK QOLFSTAILG FALSEANGFL CLMVAFLILF AM

11AA\_SEQUENCE 1.0  
P1:RKTOS2 - ribulose-bisphosphate carboxylase (EC 4.1.1.39) small chain 2  
Precursor - tomato  
C:Species: Lycopersicon esculentum (tomato)  
C>Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 18-Jun-1999  
C:Accession: S02363; B24885; B29037; T07157  
R:Sugita, M.; Manzara, T.; Pichersky, E.; Cashmore, A.; Grussem, W.  
Mol. Gen. Genet. 209, 247-256, 1987  
A:Title: Genomic organization, sequence analysis and expression of all five  
genes encoding the small subunit of ribulose-1,5-bisphosphate  
carboxylase/oxygenase from tomato.  
A:Reference number: S01107; M01D:86038372  
A:Accession: S02363  
A:Molecule type: DNA  
A:Residues: 1-180 <SUG>  
A:Cross-references: EMBL:X05983  
A>Note: the authors translated the codon TCA for residue 32 as Thr and ACT for  
residue 34 as Ser  
R:Pichersky, E.; Bernatzky, R.; Tanksley, S.D.; Cashmore, A.R.  
Proc. Natl. Acad. Sci. U.S.A. 83, 3880-3884, 1986  
A:Title: Evidence for selection as a mechanism in the concerted evolution of  
Lycopersicon esculentum (tomato) genes encoding the small subunit of  
ribulose-1,5-bisphosphate carboxylase/oxygenase.  
A:Reference number: A24885; M01D:86233336  
A:Accession: B24885  
A:Molecule type: mRNA  
A:Residues: 1-180 <PIC>  
A:Cross-references: EMBL:M13543; NID:q170497; PIDN:AAA4189.1; PID:q170498  
A>Note: the authors translated the codon GTC for residue 5 as Ala  
R:McNIGHT, T.D.; Alexander, D.C.; Babcock, M.S.; Simpson, R.B.  
Gene 48, 23-32, 1986  
A:Title: Nucleotide sequence and molecular evolution of two tomato genes  
encoding the small subunit of ribulose-1,5-bisphosphate carboxylase.  
A:Reference number: A29037; M01D:87163513  
A:Accession: B29037  
A:Molecule type: mRNA  
A:Residues: 1-86, 'V', 88-180 <MCK>  
A:Cross-references: EMBL:M15236; NID:q170503; PIDN:AAA4192.1; PID:q170504  
R:Manzara, T.  
submitted to the EMBL Data Library, August 1989  
A:Reference number: 215964  
A:Accession: T07157  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-180 <MAN>  
A:Cross-references: EMBL:X05983; NID:q19331; PIDN:CAA29401.2; PID:el389536  
A:Experimental source: cultivar VENT LA1221 cherry line  
C:Genetics:  
A:Gene: rbcS-2  
A:Map position: 3  
A:Insertions: 59/3; 104/3; 122/2  
C:Superfamily: ribulose-bisphosphate carboxylase small chain  
C:Keywords: Calvin cycle; carbon dioxide fixation; carbon-carbon lyase;  
carboxy-lyase; chloroplast; monooxygenase; photorespiration  
F:1-57/Domain: transit peptide (chloroplast) #status predicted <TMP>  
F:58-180/Product: ribulose-bisphosphate carboxylase small chain 2 #status  
Predicted <MAT>

RKTOS2 Length: 180 February 11, 2000 15:51 Type: P Check: 4535 ..

1 MASSIVSSAA VATRSNVTQA SWAPFTGLK SSATFPYTK ONLDTISAS  
51 NGRVSCMOV WPIINKKYE TLYSLPDLSD EQLLSEIEVL LKNGVPCLE  
101 FETHEGFYR ENNKSPGYD GRYWTMKLP MFCCTDATOV LAEVOEAKKA  
151 YPOAWVRILG FDMVRQVOCI SFIAKPEGY

11AA\_SEQUENCE 1.0  
P1:RKTOS1 - ribulose-bisphosphate carboxylase (EC 4.1.1.39) small chain 1  
Precursor - tomato  
C:Species: Lycopersicon esculentum (tomato)  
C>Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 18-Jun-1999  
C:Accession: S02364; A24885; A29037; S21583  
R:Sugita, M.; Manzara, T.; Pichersky, E.; Cashmore, A.; Grussem, W.  
Mol. Gen. Genet. 209, 247-256, 1987  
A:Title: Genomic organization, sequence analysis and expression of all five  
genes encoding the small subunit of ribulose-1,5-bisphosphate  
carboxylase/oxygenase from tomato.  
A:Reference number: S01107; M01D:86038372  
A:Accession: S02364  
A:Molecule type: DNA  
A:Residues: 1-181 <SUG>  
A:Cross-references: EMBL:X05982; NID:q19326; PIDN:CAA29400.1; PID:q295814  
A>Note: the authors translated the codon GTC for residue 49 as Ile  
R:Pichersky, E.; Bernatzky, R.; Tanksley, S.D.; Cashmore, A.R.  
Proc. Natl. Acad. Sci. U.S.A. 83, 3880-3884, 1986  
A:Title: Evidence for selection as a mechanism in the concerted evolution of  
Lycopersicon esculentum (tomato) genes encoding the small subunit of  
ribulose-1,5-bisphosphate carboxylase/oxygenase.  
A:Reference number: A24885; M01D:86233336  
A:Accession: A24885  
A:Molecule type: DNA  
A:Residues: 1-181 <PIC>  
A:Cross-references: EMBL:M13542; NID:q170495; PIDN:AAA4188.1; PID:q170496  
R:McNIGHT, T.D.; Alexander, D.C.; Babcock, M.S.; Simpson, R.B.  
Gene 48, 23-32, 1986  
A:Title: Nucleotide sequence and molecular evolution of two tomato genes  
encoding the small subunit of ribulose-1,5-bisphosphate carboxylase.  
A:Reference number: A29037; M01D:87163513  
A:Accession: A29037  
A:Molecule type: mRNA  
A:Residues: 1-181 <MCK>  
A:Cross-references: EMBL:M15235  
R:Manzara, T.; Carrasco, P.; Grussem, W.  
submitted to the EMBL Data Library, April 1992  
A:Description: Developmental and organ-specific changes in DNA-protein  
interactions in the tomato rbcS1, rbcS2 and rbcS3A promoter regions.  
A:Reference number: S21582  
A:Accession: S21583  
A:Molecule type: DNA  
A:Residues: 1-20 <MAN>  
A:Cross-references: EMBL:X66068; NID:q22621; PIDN:CAA46868.1; PID:q22622  
C:Genetics:  
A:Gene: rbcS-1  
A:Map position: 2  
A:Insertions: 60/3; 105/3  
C:Superfamily: ribulose-bisphosphate carboxylase small chain  
C:Keywords: Calvin cycle; carbon dioxide fixation; carbon-carbon lyase;  
carboxy-lyase; chloroplast; monooxygenase; photorespiration  
F:1-58/Domain: transit peptide (chloroplast) #status predicted <TMP>  
F:59-181/Product: ribulose-bisphosphate carboxylase small chain 1 #status  
Predicted <MAT>

RKTOS1 Length: 181 February 11, 2000 15:51 Type: P Check: 5240 ..

1 MASSIVSSAA AATRSNVAQA SWAPFTGK SAASEPVYTK NNVVDITSLA  
51 SNGRVRCMO VWPINMKRY ETLSTYPLDS DEQLISEIEY LKNGVPCLE  
101 EFERGERFYR RENNSSPGY DGRYWTMKL PMFGCTDATO VLAEOEAKK

```

151  AYPQAWIRII GEDNVRQVOC ISFIAYKPEG F

!!AA_SEQUENCE 1.0
P1:RKPOS8 - ribulose-bisphosphate carboxylase (EC 4.1.1.39) precursor small
chain rbcS-c - potato
C:Species: Solanum tuberosum (potato)
C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 18-Jun-1999
C:Accession: A31083
R:Wolter, F.P.; Fritsch, C.C.; Willmitzer, L.; Schell, J.; Schreier, P.H.
Proc. Natl. Acad. Sci. U.S.A. 85, 846-850, 1988
A:Title: rbcS genes in Solanum tuberosum: conservation of transit peptide and
exon shuffling during evolution.
A:Reference number: A31083; MUID:88124937
A:Accession: A31083
A:Molecule type: mRNA
A:Residues: 1-181 <WOL>
A:Cross-References: GB:J03613; NID:g169556; PIDN:AAA33838.1; PID:g169557
A:Experimental source: strain HH1201/7
C:Comment: Ribulose-bisphosphate carboxylase, a major component of leaf
protein, is also a monooxygenase; it catalyzes the carboxylation of D-ribulose
1,5-bisphosphate (the primary event in photosynthetic carbon dioxide fixation)
as well as the oxidative fragmentation of the pentose substrate in the
photorespiration process. These reactions occur simultaneously and in
competition at the same active site.
C:Comment: Each active molecule contains eight large chains, synthesized on the
chloroplast ribosomes and containing the active site, and eight small chains,
the precursors of which are synthesized on cytoplasmic ribosomes and converted
to mature small chains during or immediately after transport into the
chloroplast.
C:Comment: This protein is coded by one member of a small multigene family.
C:Genetics:
A:Gene: rbcS-c
A:Superfamily: ribulose-bisphosphate carboxylase small chain
C:Keywords: Calvin cycle; carbon dioxide fixation; carbon-carbon lyase;
carbon-oxygen lyase; carboxy-lyase; chloroplast; monooxygenase; photorespiration
F:1-57/Domain: transit peptide (chloroplast) #status predicted <TMP>
F:58-181/Product: ribulose-bisphosphate carboxylase small chain rbcS-c #status
predicted <MAT>
RKPOS8 Length: 181 February 11, 2000 15:51 Type: P Check: 4699 ..

1  MASSIVSSAA VATRSNVAQA SMVAPFTGLK SAASFPVYRK NNNVDITSLA
51  SNGGRVRCMQ WMPPIRMKRY ETLSTLPDLT DEQLKEVEY LKNGWVPC
101  EFETHEGFYV REHNSSPGY DGRWTMWKL PMFGCTDGTQ VLAEOEAKN
151  AYPQAWIRII GEDNVRQVOC ISFIAYKPEG Y

!!AA_SEQUENCE 1.0
P1:RKPOS1 - ribulose-bisphosphate carboxylase (EC 4.1.1.39) precursor small
chain rbcS-1 - potato
C:Species: Solanum tuberosum (potato)
C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 18-Jun-1999
C:Accession: B31083
R:Wolter, F.P.; Fritsch, C.C.; Willmitzer, L.; Schell, J.; Schreier, P.H.
Proc. Natl. Acad. Sci. U.S.A. 85, 846-850, 1988
A:Title: rbcS genes in Solanum tuberosum: conservation of transit peptide and
exon shuffling during evolution.
A:Reference number: A31083; MUID:88124937
A:Accession: B31083
A:Molecule type: DNA
A:Residues: 1-181 <WOL>
R:Fritsch, C.C.; Wolter, F.P.; Schenkemeyer, V.; Hergel, T.; Schreier, P.H.
submitted to the EMBL Data Library, December 1992
A:Reference number: S31494
A:Accession: S31494
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-181 <FRI>
A:Cross-References: EMBL:X69759; NID:g21562; PIDN:CAA9413.1; PID:g21563
C:Genetics:

```

---

```

A:Gene: rbcS-1
A:Introns: 59/3; 102/3
C:Superfamily: ribulose-bisphosphate carboxylase small chain
C:Keywords: Calvin cycle; carbon dioxide fixation; carbon-carbon lyase;
carbon-oxygen lyase; carboxy-lyase; chloroplast; monooxygenase; photorespiration
F:1-57/Domain: transit peptide (chloroplast) #status predicted <TMP>
F:58-181/Product: ribulose-bisphosphate carboxylase small chain rbcS-1 #status
predicted <MAT>
RKPOS1 Length: 181 February 11, 2000 15:51 Type: P Check: 4863 ..

1  MASSIVSSAA VATRNVTOA GSMIAPEFGL KSNATFVSR KONLDITSLA
51  SNGGRVRCMQ WMPPIRMKRY ETLSTLPDLT DEQLKEVEY LKNGWVPC
101  EFETHEGFYV REHNSSPGY DGRWTMWKL PMFGCTDGTQ VLAEOEACK
151  SYPQAWIRII GEDNVRQVOC ISFIAYKPEG Y

!!AA_SEQUENCE 1.0
P1:RKPS8 - ribulose-bisphosphate carboxylase (EC 4.1.1.39) small chain (ssu8)
precursor - garden petunia
C:Species: Petunia x hybrida (garden petunia)
C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 18-Jun-1999
C:Accession: A24917
R:Turner, N.E.; Clark, W.G.; Tabor, G.J.; Hironaka, C.M.; Fraley, R.T.; Shah,
D.N.
Nucleic Acids Res. 14, 3325-3342, 1986
A:Title: The genes encoding the small subunit of ribulose-1,5-bisphosphate
carboxylase are expressed differentially in petunia leaves.
A:Reference number: A93619; MUID:86205237
A:Accession: A24917
A:Molecule type: DNA
A:Residues: 1-180 <TURN>
A:Cross-References: GB:X03820; NID:g20490; PIDN:CAA2444.1; PID:g20491
C:Genetics:
A:Gene: rbcS
A:Introns: 59/3; 104/3; 122/3
C:Superfamily: ribulose-bisphosphate carboxylase small chain
C:Keywords: Calvin cycle; carbon dioxide fixation; carbon-carbon lyase;
carbon-oxygen lyase; carboxy-lyase; chloroplast; monooxygenase; photorespiration
F:1-57/Domain: transit peptide (chloroplast) #status predicted <TMP>
F:58-180/Product: ribulose-bisphosphate carboxylase small chain (ssu8) #status
predicted <MAT>
RKPS8 Length: 180 February 11, 2000 15:51 Type: P Check: 2967 ..

1  MASSIVSSAA VATRNVQA SMVAPFNGLK SAVSFPVSK ONLDITSLAS
51  NGRVRCMQW WPPYGRKRYE TLTSLPDLT DEQLKEVEY LKNGWVPC
101  FETHEGFYV EYHASPYYD GRWTMWKLP MFCCTDGTQ VLDELAKKA
151  YPNAMIRIIG FDNVRQVOCI SFIAKPPGF

!!AA_SEQUENCE 1.0
P1:F69398 - dihydroxyaciphthoic acid synthase (menB) homolog - Archaeoglobus
fulgidus
C:Species: Archaeoglobus fulgidus
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: F69398
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum,
K.A.; Dodson, R.J.; Gaitan, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.;
Kerlavage, A.R.; Graham, D.E.; Kyrillos, N.C.; Fleischmann, R.D.; Quackenbush,
J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kinkness, E.F.; Dougherty, B.A.;
McKenry, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeill, L.K.;
Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.;
McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artlach, P.; Kaine, B.P.;
Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;
Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic

```

sulfate-reducing archaeon *Archaeoglobus fulgidus*.  
 A:Reference number: A69250; MUID:98049343  
 C:Accession: F09398  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-277 <KLE>  
 A:Cross-references: GB:AE001022; GB:AE000782; NID:g2689345; PIDN:AA90054.1;  
 PID:g2649394; TIGR:AF1191  
 C:Superfamily: naphthate synthase; enoyl-CoA hydratase homology  
 F:29-192/Domain: enoyl-CoA hydratase homology <ECH>  
 F69398 Length: 277 February 11, 2000 15:51 Type: P Check: 4436 ..

1 MGFLFODII YEKEGRVAKI TINREKINA CTPVTVLEIS KAFIDAWTDR  
 51 KIGVVVETGA GDKAFVCGD OSIRLDGYS YSSELEGTI AALPLEVGMQ  
 101 IYTFILRHIP KPIAVRNGY AVGGHVMQV NCDLSIASEK AAFGQAGPRV  
 151 GSFDDGFGTG ELWRNVGMKR AKELWFLCEL YTAELALKMG LVNAVVPHEK  
 201 LDEEVKMS ELLEKSPTAL KMLKYAFIAD TEGLAGITEL GVGGLSLYVG  
 251 TEESLEGRNA FMEKRKDFW KEVEGGD

11AA-SEQUENCE 1.0  
 PI:F71428 hypothetical protein - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 A:Variety: Columbia  
 C>Date: 10-Sep-1998 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: F71428  
 A:Residues: 1-174 <BNV>  
 A:Cross-references: GB:297340; NID:g2244950; PIDN:CAB10400.1; PID:e327497;  
 PID:g2244979  
 C:Genetics:  
 A:Map position: 4COP9-4G3845  
 C:Superfamily: naphthate synthase; enoyl-CoA hydratase homology  
 F:31-180/Domain: enoyl-CoA hydratase homology <ECH>  
 F71428 Length: 244 February 11, 2000 15:51 Type: P Check: 9250 ..

1 MDOTVSENLI QVKESGIA VITINPKSL NSLTRAMVD LAAAFKDMDS  
 51 DESVQVVIET GSGRSCSGV DLTAASEVR GDVNDPEPDP VYOMERLARP  
 101 IIGAINGFAT TAGFELALAC DILVASGAK FMDTHARFGI FPSMGISOKL  
 151 SRIIGANKAR EVSLTSMPLT ADVAGKLGFEV NHVVEGEGAL KKAREIAEAI  
 201 IKNEQGVLR IKSVINDLK LDGHALTLE KYKSHAIENL LINDL

11AA-SEQUENCE 1.0  
 PI:A48681 - placental thrombin inhibitor - human

N:Alternate names: cytoplasmic antiprotease; intracellular serine proteinase inhibitor, 38k  
 C:Species: Homo sapiens (man)  
 C>Date: 07-Apr-1994 #sequence\_revision 07-Jul-1995 #text\_change 18-Jun-1999  
 C:Accession: A48681; A54352; A46672; B46672; C46672; S35750  
 R:Conchlin, P.; Sun, J.; Cerruti, L.; Salem, H.H.; Bird, P.  
 Proc. Natl. Acad. Sci. U.S.A. 90, 9417-9421, 1993  
 A:Title: Cloning and molecular characterization of a human intracellular serine proteinase inhibitor.  
 A:Reference number: A48681; MUID:94022386  
 A:Accession: A48681  
 A:Molecule type: mRNA  
 A:Residues: 1-376 <COO>  
 A:Cross-references: GB:222658; NID:g297411; PIDN:CA80373.1; PID:g297412  
 A:Experimental source: Placenta  
 A:Note: authors translated the codon CAA for residue 198 as Gly  
 R:Morgenstern, K.A.; Sprecher, C.; Holth, L.; Foster, D.; Grant, F.J.; Ching, A.; Kiesel, W.  
 Biochemistry 33, 3432-3441, 1994  
 A:Title: Complementary DNA cloning and kinetic characterization of a novel intracellular serine proteinase inhibitor: mechanism of action with trypsin and factor Xa as model proteinases.  
 A:Reference number: A54352; MUID:94183847  
 A:Accession: A54352  
 A:Molecule type: mRNA  
 A:Residues: 1-174; 'E', 176-361; 'S', 363-376 <MOR>  
 A:Cross-references: GB:569272; NID:g546087; PIDN:AA830320.1; PID:g546088  
 A:Experimental source: Placenta  
 A:Note: sequence extracted from NCBI backbone (NCBIN:145231, NCBI:P:145232)  
 R:Conchlin, P.B.; Tetaz, T.; Salem, H.H.  
 J Biol Chem 268, 9541-9547, 1993  
 A:Title: Identification and purification of a novel serine proteinase inhibitor.  
 A:Reference number: A46672; MUID:93252826  
 A:Accession: A46672  
 A:Molecule type: Protein  
 A:Residues: 47-60; 63-81; 91-98 <CO2>  
 A:Experimental source: Placenta, leukemic cell line K562  
 A:Note: sequence modified after extraction from NCBI backbone  
 C:Genetics:  
 A:Gene: GDB:PI6  
 A:Cross-references: GDB:252025; OMIM:173321  
 A:Map position: 6p25-6p24.3  
 C:Superfamily: antithrombin III  
 C:Keywords: blocked amino end; cytosol; serine proteinase inhibitor  
 F:341/inhibitory site: Arg (thrombin) #status predicted  
 A48681 Length: 376 February 11, 2000 15:51 Type: P Check: 300 ..

1 MDVLAENGT FALNLTG KDNKNVFRS PMSKSCALAM VYMGAKGNTA  
 51 AQMAQLISFN KSGGGGDH QFOSLITEVN KTGYLYLRV ANRLEGEKSC  
 101 DFLSFRDSC QKFOAEMEE LDFISAVES RKHINTWVAE KTEGIAEELL  
 151 SPGSVDPLR LVVNAVYFR GMDGQDFKE NTEERLEKVS KNEKPVQMM  
 201 FKSTFKKTY IGEIFTQILV LPYVKELMM IIMPDETTO LRYVEKELRY  
 251 EKVEVTRFD MMDDEVEVS LPRFKLESY DMESVLNLNG MDAFLGLGA  
 301 DESGMSOTDL SLKVKYKSF VEYNEGCTEA AAATRAIMM RCRAPFRRC  
 351 ADHPLEFTIO HRKTNGLILFC GRESSP

11AA-SEQUENCE 1.0  
 PI:W17A - alpha-amylase inhibitor CII - wheat  
 C:Species: Triticum aestivum (common wheat)  
 C>Date: 18-Aug-1982 #sequence\_revision 18-Aug-1982 #text\_change 04-Sep-1998  
 C:Accession: A01322  
 R:Kashlan, N.; Richardson, M.  
 Phytochemistry 20, 1781-1784, 1981  
 A:Title: The complete amino acid sequence of a major wheat protein inhibitor of alpha-amylase.

```
A:Reference number: A01322
A:Accession: A01322
A:Molecule type: protein
A:Residues: 1-123 <KAS>
A:Note: half of the molecules contained 65-Ser. Very small amounts of 67-Ala,
98-Gly, 99-Pro, and 118-Val were also found.
C:Comment: This protein is one of a number of related alpha-amylase inhibitors
found in the endosperm of wheat seeds.
C:Superfamily: wheat alpha-amylase inhibitor
C:Keywords: alpha-amylase inhibitor
F:7-52,21-42,29-82,43-98,56-113/Disulfide bonds: #status predicted
MIWTA Length: 123 February 11, 2000 15:51 Type: P Check: 4689 ..

1 SGPWMCNDA TGYKVSALTG CRAWKLCY GSOVPEAVLR DCCQQLADIN
51 NEMCRGDIS SMLRAYOEL GVEGKEVLP GCRKEVMKL AASVEVCKV
101 PIPNPSGDRA GVCYGDWCAY PDV

!!AA_SEQUENCE 1.0
P1:MIWTA5 - alpha-amylase inhibitor 0.53 - wheat
C:Species: Triticum aestivum (common wheat)
C:Date: 13-Jun-1983 #sequence_revision 28-Feb-1986 #text_change 04-Sep-1998
C:Accession: A90649; A90656; A01323
R:Maeda, K.; Hase, T.; Matsubara, H.
Biochim. Biophys. Acta 743, 52-57, 1983
A:Title: Complete amino acid sequence of an alpha-amylase inhibitor in wheat
kernel
A:Reference number: A90649; MUID:83127436
A:Accession: A90649
A:Molecule type: protein
A:Residues: 1-119, YPPKA <MA>
A:Note: this sequence has been revised in reference A90656
R:Maeda, K.; Wakabayashi, S.; Matsubara, H.
Biochim. Biophys. Acta 828, 213-221, 1985
A:Title: Complete amino acid sequence of an alpha-amylase inhibitor in wheat
kernel (0.19-inhibitor)
A:Reference number: A90656; MUID:85175148
A:Accession: A90656
A:Molecule type: protein
A:Residues: 1-124 <MA>
R:Maeda, K.; Wakabayashi, S.; Matsubara, H.
J. Biochem. 94, 865-870, 1983
A:Title: Disulfide bridges in an alpha-amylase inhibitor from wheat kernel.
A:Reference number: A91976; MUID:84061717
A:Accession: A91976
A:Contents: annotation: disulfide bonds
C:Comment: This protein is a dimer of identical chains.
C:Superfamily: wheat alpha-amylase inhibitor
C:Keywords: alpha-amylase inhibitor
F:20-41,42-99,54-115/Disulfide bonds: #status predicted
F:28-83/Disulfide bonds: #status experimental
MIWTA5 Length: 124 February 11, 2000 15:51 Type: P Check: 4073 ..

1 SGPWMCYPCQ AFQVPALPGC RPLIKQCNQ SOVPEAVLRD CCQQLADISE
51 WPRGALYSM LDSMKREHG SEGOAGTGAF PSGRREVVL TAASITAVCR
101 LPIVDASGD GAYVCKDVAA YPDA

!!AA_SEQUENCE 1.0
P1:MIWTA1 - alpha-amylase inhibitor 0.19 - wheat
C:Species: Triticum aestivum (common wheat)
C:Date: 28-Feb-1986 #sequence_revision 28-Feb-1986 #text_change 30-Sep-1998
C:Accession: A01324
R:Maeda, K.; Wakabayashi, S.; Matsubara, H.
Biochim. Biophys. Acta 828, 213-221, 1985
A:Title: Complete amino acid sequence of an alpha-amylase inhibitor in wheat
kernel (0.19-inhibitor)
A:Reference number: A90656; MUID:85175148
A:Accession: A01324
A:Molecule type: protein

A:Residues: 1-124 <MA>
R:Okuda, M.; Satoh, T.; Sakurai, N.; Shibuya, K.; Kaji, H.; Samejima, T.
J. Biochem. 122, 918-926, 1997
A:Title: Overexpression in Escherichia coli of chemically synthesized gene for
active 0.19 alpha-amylase inhibitor from wheat kernel.
A:Reference number: JC5705; MUID:98104043
A:Accession: JC5705
A:Contents: annotation
C:Comment: For the sequence from a synthetic DNA used for artificial protein
production, see PIR:JC5705.
C:Complex: homodimer
C:Superfamily: wheat alpha-amylase inhibitor
C:Keywords: alpha-amylase inhibitor; homodimer
F:8-52,20-41,28-83,42-99,54-115/Disulfide bonds: #status predicted
F:15-52,20-41,28-83,42-99,54-115/Disulfide bonds: #status predicted
MIWTA1 Length: 124 February 11, 2000 15:51 Type: P Check: 3320 ..

1 SGPWMCYPCQ AFQVPALPGC RPLIKQCNQ SOVPEAVLRD CCQQLADISE
51 WPRGALYSM LDSMKREHGA QEGOAGTGAF PSGRREVVL TAASITAVCR
101 LPIVDASGD GAYVCKDVAA YPDA

!!AA_SEQUENCE 1.0
P1:TVHUF1 - transforming protein fos - human
C:Species: Homo sapiens (man)
C:Date: 03-Aug-1984 #sequence_revision 03-Aug-1984 #text_change 24-Sep-1999
C:Accession: A01342; I58359; S09375
R:Van Straaten, F.; Muller, R.; Curran, T.; Van Beveren, C.; Verma, I.M.
Proc. Natl. Acad. Sci. U.S.A. 80, 3183-3187, 1983
A:Title: Complete nucleotide sequence of a human c-onc gene: deduced amino acid
sequence of the human c-fos protein
A:Reference number: A01342; MUID:83221560
A:Accession: A01342
A:Molecule type: DNA
A:Residues: 1-360 <VAN>
A:Cross-references: GB:V01512; NID:929903; PIDN:CAA24756.1; PID:929904
R:Nakabeppu, Y.; Nathans, D.
EMBO J. 8, 3833-3841, 1989
A:Title: The basic region of Fos mediates specific DNA binding.
A:Reference number: S09374; MUID:90059986
A:Accession: S09374
A:Contents: annotation
A:Note: engineered chimeric forms including residues 121-207
R:Boux, P.; Verrier, B.; Klein, B.; Niccolino, M.; Marty, L.; Alexandre, C.;
Piechaczyk, M.
Oncogene 6, 2155-2160, 1991
A:Title: Retrovirus-mediated gene transfer of a human c-fos cDNA into mouse
bone marrow stromal cells.
A:Reference number: I58359; MUID:92050815
A:Accession: I58359
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-6 <RSS>
A:Cross-references: GB:S65138; NID:9238776; PIDN:AAB20306.1; PID:9238778
C:Genetics:
A:Gene: GDB:FOS
A:Cross-references: GDB:119917; OMIM:164810
A:Map position: 14q24.3-14q24.3
A:Introns: 47/3; 131/3; 167/3
C:Superfamily: fos transforming protein; fos/jun DNA-binding domain homology
C:Keywords: DNA binding; heterodimer; leucine zipper
F:132-172/Domain: fos/jun DNA-binding domain homology <FUD>
F:165-193/Region: leucine zipper motif
TVHUF1 Length: 380 February 11, 2000 15:51 Type: P Check: 4677 ..

1 MPEFSGNADY EASSRCSA SPAGDSLSTY HSPADSFSM GSPVNAQDFC
51 TDLAVSSANF IPTVTAISTS PDLQWLQPA LVSSVAPSGT RAPHFEGVPA
101 PSNGATSRAG VYKTMIGRA QSIGRGKYV QLSPEEEKR RIRERKXMA
151 AKCRNRRE LTDTLQAEIT QLEDEKSAIQ TEIANLKEK EKLEFIILAH
```



```

201 RPACKIPDDL GFPEKSVAS LDLTGLPEV ATPSEBAFT LPLNDPEPK
251 PSVEPVKIS SMELKTEPPD DELFPASSRP SGSEFARSVP DMDLSGSFYA
301 ADMEPLHSGS LGMGPATTEL EPLCTPVYTC TPSCFAYTSS FVFTYPEADS
351 FPCCAAHARK GSSSEPPSSD SLSSPTLLAL

!!AA_SEQUENCE 1.0
PI:TVFV4 - transforming protein fos - chicken
C:Species: Gallus gallus (chicken)
C>Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 18-Jun-1999
C:Accession: A28368; A34073
R:Moelders, H.; Jenwein, T.; Adamkiewicz, J.; Mueller, R.
Oncogene 1, 377-385, 1987
A:Title: Isolation and structural analysis of a biologically active chicken
c-fos cDNA: identification of evolutionarily conserved domains in fos protein.
A:Reference number: A28368; MUID:88262231
A:Accession: A28368
A:Molecule type: mRNA
A:Residues: 1-367 <MOB>
A:Cross-references: GB:M37000; NID:9211469; PIDN:AAA48670.1; PID:9211470
K:Fujimura, K.T.; Ashida, K.; Nishina, H.; Iba, H.; Miyajima, N.; Nishizawa,
M.; Kawai, S. 4012-4018, 1987
J:Virol. 61, 4012-4018, 1987
A:Title: The chicken c-fos gene: cloning and nucleotide sequence analysis.
A:Reference number: A34073; MUID:88062957
A:Accession: A34073
A:Molecule type: DNA
A:Residues: 1-367 <FUD>
A:Cross-references: GB:M18043; NID:9211467; PIDN:AAV76823.1; PID:9211468
A:Note: the authors translated the codon AAC for residue 90 as Asp
C:Genetics:
A:Gene: fos
A:introns: 47/3; 130/3; 166/3
C:Superfamily: fos transforming protein; fos/jun DNA-binding domain homology
C:Keywords: DNA binding; leucine zipper; transforming protein
F:331-171/Domain: fos/jun DNA-binding domain homology <FUD>
F:164-192/Region: leucine zipper motif

TVCHF5 Length: 367 February 11, 2000 15:51 Type: P Check: 7007

1 MMVQFAGEX EAPSRCSA SPAGDSLTYY PSPADSEFSM GSPVNSQDFC
51 TDLAVSANF VPTVAISTS PDLQWLVOPT LISSVAFSON RGHVGVAPAP
101 APPAAVSRA VLKAPGKQVQ SIGRKGKVGQ LSPSEEEKR IRREKMKMA
151 AKCRNRREL TDLTAEEDQ LEEKSALQA EIANLKEKE KLEFILAHR
201 PACKMEELR FSEELAAATA LDGAPSPAA AEAFALPIM TEAPVAPPK
251 EPGSGLELK AEPDELLFS AGPREASRV PDMDLPASS FYASDMEPLG
301 AGSGGELEPL CTPTVYCTPC PSTYTSIFYF TYPEADAPPS CAAAHKRGSS
351 SNEPSSDLS SPTLLAL

!!AA_SEQUENCE 1.0
PI:TVFV4 - transforming protein fos - avian retrovirus NK24
C:Species: avian retrovirus NK24
C>Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 26-May-1995
C:Accession: B34071
R:Nishizawa, M.; Goto, N.; Kawai, S.
J:Virol. 61, 3733-3740, 1987
A:Title: An avian transforming retrovirus isolated from a nephroblastoma that
carries the fos gene as the oncogene.
A:Reference number: A34071; MUID:88062920
A:Accession: B34071
A:Molecule type: DNA
A:Residues: 1-322 <NIS>
C:Genetics:
A:Gene: fos

C:Superfamily: fos transforming protein; fos/jun DNA-binding domain homology
C:Keywords: DNA binding; leucine zipper; transforming protein
F:86126/Domain: fos/jun DNA-binding domain homology <FUD>
F:119-147/Region: leucine zipper motif

TVFV4 Length: 322 February 11, 2000 15:51 Type: P Check: 9324

1 SQDFCTDLAV SSANFPTVT AISTSPDLQW LYQPTLLSSV ARSQRGHPY
51 GVPAPAPPA YSRPAVLKAP GGRGOSIGRR GKVEQLSPEE EKKRIIRER
101 NKMAAKCEN RRRLDTLQ AETDLEEK SALQAEIANT LKEKELEFI
151 LAAHRPCKM PEELFSEEL AATALDUGA PSPAAEEAF ALPLMTEAP
201 AVPPKEPSGS GLEKAEFPD ELLFSAGPRE ASRSVPMDL PGASSFYASD
251 WEPLGASGG ELEPLCTPVY TCTPCPSTYT STVFITYPEA DAFPCCAAH
301 RKGSSNEPS SDLSPTLL AL

!!AA_SEQUENCE 1.0
PI:S00755 - pleckstrin - human
N:Alternate names: p47; platelet/leukocyte C kinase substrate (pleckstrin)
C:Species: Homo sapiens (man)
C>Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 16-Jul-1999
C:Accession: S00755; A45762
R:Tyers, M.; Rachubinski, R.A.; Stewart, M.I.; Varrichio, A.M.; Short, R.G.L.;
Haslam, R.J.; Harley, C.B.
Nature 333, 470-473, 1988
A:Title: Molecular cloning and expression of the major protein kinase C
substrate of platelets
A:Reference number: S00755; MUID:88232910
A:Accession: S00755
A:Molecule type: mRNA
A:Residues: 1-350 <TYD>
A:Cross-references: EMBL:X07743; NID:935517; PIDN:CAA30564.1; PID:935518
R:Tyers, M.; Haslam, R.J.; Rachubinski, R.A.; Harley, C.B.
J:Cell. Biochem. 40, 133-143, 1989
A:Title: Molecular analysis of pleckstrin: the major protein kinase C substrate
of platelets.
A:Reference number: A45762; MUID:89359547
A:Accession: A45762
A:Molecule type: mRNA
A:Residues: 1-350 <TYD>
A:Cross-references: GB:X07743; NID:935517; PIDN:CAA30564.1; PID:935518
A:Note: 92-Arg was also found
C:Superfamily: pleckstrin; pleckstrin repeat homology
C:Keywords: phosphoprotein; signal transduction
F:3-99/Domain: pleckstrin repeat homology <PLK1>
F:243-345/Domain: pleckstrin repeat homology <PLK2>

S00755 Length: 350 February 11, 2000 15:51 Type: P Check: 1536

1 MEPRRIEGY LVKGSVENT WKPMWVYLE DGIEFYKKKS DNSPKMIPL
51 KGSLTSPQC DCKRMFVFK ITTKQODHF FQAALEERD AWARDINKAI
101 KCIEGQKFA RKSTRSIRL PETIDGALV LEMKDTERRI KELULEDRK
151 IFNCFIENG VYDMLVSNQS VRNRQECIMI ASLLNEGYL QPAGDMKSA
201 VDGTAENPFL DNPDAFYFP DSGFCEENS SDDVILKEE FRGVIKQGC
251 LKQGHRRKN WKYKFIKRE DPALVHYDP AGADEPLGAI HLRCQVTSV
301 ESNNGKXSE EENLEFIITA DEVHYFLQAA TKERTENIK AIQWASRTGK

!!AA_SEQUENCE 1.0
PI:HLH0G - class II histocompatibility antigen-associated gamma chain - human
N:Alternate names: CD74 antigen; HLA-DR antigen-associated invariant chain; MHC
class II-associated invariant chain
C:Species: Homo sapiens (man)

```

C>Date: 04-Dec-1986 #sequence\_revision 26-Jan-1996 #text\_change 22-Jun-1999  
 C/Accession: A93981; B93981; A93602; A94103; S28903; S07182; A33234; A02243;  
 A27551; A30060; S46255  
 R/Claesson, L.; Larhammar, D.; Rask, L.; Peterson, P.A.  
 Proc. Natl. Acad. Sci. U.S.A. 80, 7395-7399, 1983  
 A/Title: cDNA clone for the human invariant gamma chain of class II  
 histocompatibility antigens and its implications for the protein structure.  
 A/Reference number: A93981; MUID:84170234  
 A/Accession: A93981  
 A/Molecule type: mRNA  
 A/Residues: 1-216 <CL1>  
 A/Cross-references: GB:K01144; NID:9188469; PIDN:AAA36304.1; PID:9188470  
 A/Accession: B93981  
 A/Molecule type: protein  
 A/Residues: 1,'XXXXX',7,'XXXXXX',14,'X',16-17 <CL2>  
 A/Note: radiochemical amino acid sequencing after cell-free translation  
 R/Kudo, U.; Chao, L.Y.; Natani, F.; Saunders, G.F.  
 Nucleic Acids Res. 13, 8827-8841, 1985  
 A/Title: Structure of the human gene encoding the invariant gamma-chain of  
 class II histocompatibility antigens.  
 A/Reference number: A93602; MUID:86093681  
 A/Accession: A93602  
 A/Molecule type: DNA  
 A/Residues: 'MHRRRSRSCREDQKPV',1-150,'T',152-216 <KUD>  
 A/Note: the authors translated the codon CAG for residue 109 as Gly, AAG for  
 residue 143 as Gln, and ACA for residue 151 as Arg  
 R/O'Sullivan, D.M.; Larhammar, D.; Wilson, M.C.; Peterson, P.A.; Quaranta, V.  
 Proc. Natl. Acad. Sci. U.S.A. 83, 4484-4488, 1986  
 A/Title: Structure of the human Ia-associated invariant (gamma)-chain gene:  
 identification of 5' sequences shared with major histocompatibility complex  
 class II genes.  
 A/Reference number: A94103; MUID:86233451  
 A/Accession: A94103  
 A/Molecule type: DNA  
 A/Residues: 1-216 <OSU>  
 R/Riberdy, J.M.; Newcomb, J.R.; Surman, M.J.; Barbosa, J.A.; Cresswell, P.  
 Nature 360, 474-477, 1992  
 A/Title: HLA-DR molecules from an antigen-processing mutant cell line are  
 associated with invariant chain peptides.  
 A/Reference number: S28903; MUID:93078879  
 A/Accession: S28903  
 A/Status: preliminary  
 A/Molecule type: protein  
 A/Residues: 81-104 <RIB>  
 R/Strubin, M.; Mach, B.; Long, E.O.  
 EMBO J. 3, 869-872, 1984  
 A/Title: The complete sequence of the mRNA for the HLA-DR-associated invariant  
 chain reveals a polypeptide with an unusual transmembrane polarity.  
 A/Reference number: S07182; MUID:84207945  
 A/Accession: S07182  
 A/Molecule type: mRNA  
 A/Residues: 'MHRRRSRSCREDQKPV',1-216 <STR>  
 A/Cross-references: EMBL:X00457; NID:932130; PIDN:CAA25192.1; PID:932131  
 A/Accession: A33234  
 A/Molecule type: DNA  
 A/Residues: 'MHRRRSRSCREDQKPV',1-4 <STR2>  
 R/Katsuma, N.; Kakegawa, H.; Matsunaga, Y.; Saitara, T.  
 FEBS Lett. 349, 265-269, 1994  
 A/Title: Immunological significances of invariant chain from the aspect of its  
 structural homology with the cystatin family.  
 A/Reference number: S46255; MUID:94326933  
 A/Contents: annotation  
 A/Note: some conclusions in this reference are based on the assumption, with no  
 experimental evidence, that an upstream initiator is used  
 C/Comment: Class II antigens are associated with the invariant gamma chain  
 during intracellular transport. In the Golgi apparatus, they are glycosylated  
 and dissociate from the gamma chain. Some gamma chains become independently  
 integrated into the plasma membrane.  
 C/Comment: Both cell-free translation experiments and the consensus sequence  
 for the initiation of translation (Kozak) suggest that Met-1 is the initiator  
 codon. Some authors have shown the translation of an additional sixteen  
 residues upstream of the initiator, but this region is not likely to expressed.  
 C/Genetics:

A/Gene: GDB:CD74; DMLAG  
 A/Cross-references: GDB:119846; OMIM:142790  
 A/Map position: 5q31-5q33  
 A/Intons: 26/2; 84/1; 110/3; 131/3; 163/3; 193/1; 214/1  
 C/Superfamily: type II histocompatibility antigen-associated gamma chain;  
 thyroglobulin type I repeat homology  
 C/Keywords: glycoprotein; transmembrane protein  
 F:1-30/Domain: intracellular #status predicted <INT>  
 F:31-56/Domain: transmembrane #status predicted <TM>  
 F:57-216/Domain: extracellular #status predicted <EXT>  
 F:114,120/Binding site: carbohydrate (asn) (covalent) #status predicted

HMUG Length: 216 February 11, 2000 15:51 Type: P Check: 908

1 MDOQRDLISN NEQPLMGRR PGAPESCSR GALYGFSL VTLLAGCAT  
 51 TATFLIQOQG RLDKLTIVSQ NLOLENIARK LRPPKPVSK MRATPPLMQ  
 101 ALPMGALPQG PQONATRYGN MTEDEVNHL QNADPLKVP PLKSPFENT  
 151 RHKNITWETI DWKVESNMWH HWLFEKSRH SLEQKPTDAP PRESLEEDP  
 201 SSGIGVTKOD LGPVPV

11AA\_SEQUENCE 1.0  
 P1:R3LV8 - ribosomal protein S8 - liverwort (Marchantia polymorpha) chloroplast  
 C/Species: Chloroplast Marchantia polymorpha  
 C/Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 22-Jun-1999  
 C/Accession: A02716; S01561  
 R/Ohyama, K.  
 submitted to the EMBL Data Library, October 1986  
 A/Reference number: A00150  
 A/Accession: A02716  
 A/Molecule type: DNA  
 A/Residues: 1-132 <OHY>  
 R/Ohyama, K.; Fukuzawa, H.; Kohchi, T.; Shirai, H.; Sano, T.; Sano, S.;  
 Umesono, K.; Shiki, Y.; Takeuchi, M.; Chang, Z.; Aota, S.; Inokuchi, H.; Ozeki,  
 H.

Nature 322, 572-574, 1986  
 A/Title: Chloroplast gene organization deduced from complete sequence of  
 liverwort Marchantia polymorpha chloroplast DNA.  
 A/Reference number: A38014  
 A/Contents: annotation; gene organization; sites; features  
 R/Fukuzawa, H.; Kohchi, T.; Sano, T.; Shirai, H.; Umesono, K.; Inokuchi, H.;  
 Ozeki, H.; Ohyama, K.  
 J. Mol. Biol. 203, 333-351, 1988  
 A/Title: Structure and organization of Marchantia polymorpha chloroplast  
 genome. III. Gene organization of the large single copy region from rbcL to  
 trnI(CAU).  
 A/Reference number: S01529; MUID:89068687  
 A/Accession: S01561  
 A/Molecule type: DNA  
 A/Residues: 1-132 <TRK>  
 A/Cross-references: GB:X04465; GB:Y00686; NID:911640; PIDN:CAA28121.1;  
 PID:911710  
 C/Genetics:  
 A/Gene: rps8  
 A/Gene: rps8  
 A/Gene: chloroplast  
 C/Superfamily: Escherichia coli ribosomal protein S8  
 C/Keywords: Chloroplast; protein biosynthesis; ribosome

R3LV8 Length: 132 February 11, 2000 15:51 Type: P Check: 6002

1 MGNDTIANMI TSIRNANLGR IKTVOYPATN ITRNIAKILF QEGFIDNFID  
 51 NKONTDILI LNLKYQKKR KSYITTLRI SRGLRYSN HKRIPKVLGG  
 101 MGVILSTR GIMTDREARQ KTIIGELLCY VM

11AA\_SEQUENCE 1.0  
 P1:S2696 - ribosomal protein S13.e. cytosolic - fission yeast  
 (Schizosaccharomyces pombe)  
 C/Species: Schizosaccharomyces pombe

C>Date: 29-Jan-1993 #sequence-revision 19-Apr-1996 #text-change 22-Jun-1999  
 C/Accession: S26296  
 R/Marks: J.J. Simanis, V.  
 Nucleic Acids Res. 20, 4094, 1992  
 A/Title: Cloning of the gene for ribosomal protein S13 from the fission yeast  
 Schizosaccharomyces pombe.  
 A/Reference number: S26296; MUID:92375702  
 A/Accession: S26296  
 A/Molecule type: DNA  
 A/Residues: 1-151 <MAR>  
 A/Cross-references: EMBL:X67030; NID:95071; PIDN:CAA47424.1; PID:95072  
 C/Genetics: 68/1: 79/1: 94/3  
 C/Insertions: rat ribosomal protein S13; eubacterial ribosomal protein S15  
 C/Superfamily: rat ribosomal protein S13; eubacterial ribosomal protein S15  
 C/Keywords: protein biosynthesis; ribosome  
 F/2-151/Product: ribosomal protein S13.e #status predicted <MAT>  
 F/82-148/Domain: eubacterial ribosomal protein S15 homology <ES15>  
 S26296 Length: 151 February 11, 2000 15:51 Type: P Check: 6968

1 MGRMSKSG IASGALPYR SPPAWCKADA DSVVEQILKF SKRGMSPSQI  
 51 GVLDSHGI POVRFITGQR IMRIKANGL APELPEDLYN LIRKAVYRK  
 101 HLERNKDKD SKFRLILIES RIHRLARYR KVGALPPTWK YESATASALV  
 151 A

1 IAA\_SEQUENCE 1.0  
 P1:AA8103 - ribosomal protein L21 precursor, chloroplast - spinach  
 C/Species: Spinacia oleracea (spinach)  
 C/Date: 10-Sep-1999 #sequence-revision 10-Sep-1999 #text-change 10-Sep-1999  
 C/Accession: AA8103; S13527; A37906  
 R/Lagarange, T.; Franca, B.; Ayala, M.; Maché, R.; Lerbs-Maché, S.  
 Mol. Cell. Biol. 13, 2614-2622, 1993  
 A/Title: Structure and expression of the nuclear gene coding for the  
 chloroplast ribosomal protein L21: developmental regulation of a housekeeping  
 gene by alternative promoters  
 A/Reference number: AA8103; MUID:93205007  
 A/Accession: AA8103  
 A/Status: Preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-256 <LAG>  
 A/Cross-references: GB:M64682; NID:9310589; PIDN:AAA74715.1; PID:9310590  
 A/Note: Sequence extracted from NCBI backbone (NCBIN:127727, NCBI:P:127728)  
 R/Martin, W.; Lagrange, T.; Li, Y.F.; Bisanz-Seyer, C.; Maché, R.  
 Curr. Genet. 18, 553-556, 1990  
 A/Title: Hypothesis for the evolutionary origin of the chloroplast ribosomal  
 protein L21 of spinach.  
 A/Reference number: S13527; MUID:91168304  
 A/Accession: S13527  
 A/Molecule type: mRNA  
 A/Residues: 1-256 <MAR>  
 A/Cross-references: EMBL:X56691; NID:921314; PIDN:CAA40019.1; PID:921315  
 R/Smooker, P.M.; Kruff, V.; Subramanian, A.R.  
 J. Biol. Chem. 265, 16699-16703, 1990  
 A/Title: A ribosomal protein is encoded in the chloroplast DNA in a lower plant  
 but in the nucleus in angiosperms. Isolation of the spinach L21 protein and  
 cDNA clone with transit and an unusual repeat sequence.  
 A/Reference number: A37906; MUID:90375547  
 A/Accession: A37906  
 A/Molecule type: mRNA  
 A/Residues: 1-256 <SMO>  
 A/Cross-references: GB:M57413; GB:M31763; NID:9170134; PIDN:AAA34041.1;  
 PID:9170135  
 C/Superfamily: Escherichia coli ribosomal protein L21  
 C/Keywords: chloroplast; protein biosynthesis; ribosome  
 F/1-53/Domain: transit peptide (chloroplast) #status predicted <TMP>  
 F/36-256/Product: ribosomal protein L21 #status predicted <MAT>  
 AA8103 Length: 256 February 11, 2000 15:51 Type: P Check: 718

1 MASALIAFSC SGLCATKLPL QNINPLILNV PPLSKPESGV VSPPSISRLS  
 51 LRVYAKRRR FOEIPPELKA EEEFQRPNN QKPQLDVLV DDFQADPEPT  
 101 PEINDLINDQF LPKGPPPPR EEIFAVVIG SRQYIVIGR WIYTORLKA  
 151 TVNDKIVLAK VLVGTRAST YIGTPIVTA AVHAAVEQL LDDKIVFXY  
 201 KKKKNTNRI GHROPITRIK ITGITGEDY PASTLEAVE AKKEAEAEAE  
 251 AEAPV

1 IAA\_SEQUENCE 1.0  
 P1:JQ1298 - annexin II type 2 - African clawed frog  
 N/Alternate names: calpactin II; lipocortin II; p36  
 C/Species: Xenopus laevis (African clawed frog)  
 C/Date: 31-Mar-1992 #sequence-revision 26-May-1994 #text-change 22-Jun-1999  
 C/Accession: JQ1298; A41002; B41002  
 R/Geis, V.; Koch, W.; Thiel, C.  
 Gene 104, 259-264, 1991  
 A/Title: Primary structure and expression of the Xenopus laevis gene encoding  
 annexin II.  
 A/Reference number: JQ1297; MUID:92009222  
 A/Accession: JQ1298  
 A/Molecule type: mRNA  
 A/Residues: 1-340 <GER>  
 A/Cross-references: GB:M60769; NID:9214532; PIDN:AAA49886.1; PID:9214533  
 A/Experimental source: Kidney  
 R/Izant, J.G.; Bryson, L.J.  
 U. Biol. Chem. 266, 18560-18566, 1991  
 A/Title: Xenopus annexin II (calpactin I) heavy chain has a distinct amino  
 terminus.  
 A/Reference number: A41002; MUID:92011609  
 A/Accession: A41002  
 A/Molecule type: mRNA  
 A/Residues: 1-222, 'HP', 225-296, 'RN', 299-340 <IZA>  
 A/Cross-references: GB:M58575; NID:9214007; PIDN:AAA49664.1; PID:9214008;  
 GB:M58576; NID:9214009; PID:9214010  
 A/Experimental source: ovary; clones A3 and C4  
 A/Accession: B41002  
 A/Molecule type: mRNA  
 A/Residues: 52-222, 'HP', 225-243, 'R', 245-257, 'S', 259-296, 'RN', 299-340 <IZ2>  
 A/Cross-references: GB:M58577  
 A/Experimental source: ovary; clones E4 and F4  
 A/Note: translation of the nucleotide sequence is not complete  
 C/Genetics:  
 A/Gene: Anx II-2  
 C/Superfamily: annexin I; annexin repeat homology  
 C/Keywords: calcium binding; duplication; endonexin fold; inflammation;  
 phospholipid binding; phosphoprotein  
 F/2-340/Product: annexin II type 2 #status predicted <MAT>  
 F/37-108/Domain: annexin repeat homology <AX1>  
 F/18-64/Region: endonexin fold #status predicted  
 F/109-188/Domain: annexin repeat homology <AX2>  
 F/120-138/Region: endonexin fold #status predicted  
 F/193-265/Domain: annexin repeat homology <AX3>  
 F/205-221/Region: endonexin fold #status predicted  
 F/269-340/Domain: annexin repeat homology <AX4>  
 F/280-296/Region: endonexin fold #status predicted  
 F/27/Binding site: phosphate (Ser) (covalent) #status predicted  
 JQ1298 Length: 340 February 11, 2000 15:51 Type: P Check: 409

1 MALIHEIGK LSLGQSSS ROSKUGSYKA ATHFDEKDA AAIETAIKTK  
 51 GVDELTINI LTRNSNQRO DIAFAFHRT KNDLPALNG ALSGNLEIYM  
 101 LGLIKTRQY DASELAKSMK GLGDEDEPLI EITCSTNKE LLDIONNRYE  
 151 LKTELEKDI MSDTSGDFRK LMAVALAGKR QEDGNVAYDE KIIDDAELEY  
 201 EAGVKKRGTD VTKWITIME RSISHLOKV ERKTSISPD IESIKKEVK

**251** GDLEMLNLN VQCIQNKKY FADLHSEMK GKCTKTKILI RMVSRCLD  
**301** MKTRDFEKK KXGSLHYFI GDDPKGDYOR ALLNLGGSD  
11AA\_SEQUENCE 1.0  
C:Species: Drosophila melanogaster  
C>Date: 30-Sep-1987 #sequence\_revision 30-sep-1987 #text\_change 22-Jun-1999  
C:Accession: A90864; A90865; A22012  
R:O.Tousa, J.E.; Baehr, W.; Martin, R.L.; Hirsh, J.; Pak, W.L.; Applebury, M.L.  
Cell 40, 839-850, 1985  
A>Title: The Drosophila rhodopsin gene encodes an opsin.  
A:Reference number: A90864; MUID:85176937  
A:Molecule type: DNA  
A:Residues: 1-373 <OTO>  
A:Cross-references: GB:X02315; NID:g158007; PIDN:AAA28733.1; PID:g158008  
R:Zuker, C.S.; Cowman, A.F.; Rubin, G.M.  
Cell 40, 851-858, 1985  
A>Title: Isolation and structure of a rhodopsin gene from Drosophila melanogaster.  
A:Reference number: A90865; MUID:85176938  
A:Molecule type: mRNA  
A:Residues: 1-373 <UW>  
C:Comment: The domains were proposed from hydrophobic indices.  
C:Comment: Some or all of the carboxyl-terminal Ser or Thr residues may be phosphorylated.  
C:Genetics:  
A:Gene: rhodopsin  
A:Cross-references: FlyBase:FBN0002940  
A:Map position: 3R66 (232B-1/3); 332/2  
A:Introns: 3/2: 190/2: 239/3: 332/2  
C:Superfamily: vertebrate rhodopsin  
C:Keywords: chromoprotein; G protein-coupled receptor; glycoprotein; photoreceptor; retinal; transmembrane protein  
F:1-49/Domain: extracellular #status predicted <EX1>  
F:50-74/Domain: transmembrane #status predicted <TM1>  
F:75-86/Domain: intracellular #status predicted <IN1>  
F:87-109/Domain: transmembrane #status predicted <TM2>  
F:110-127/Domain: extracellular #status predicted <EX2>  
F:128-153/Domain: transmembrane #status predicted <TM3>  
F:154-160/Domain: intracellular #status predicted <IN2>  
F:161-181/Domain: transmembrane #status predicted <TM4>  
F:182-215/Domain: extracellular #status predicted <EX3>  
F:216-243/Domain: transmembrane #status predicted <TM5>  
F:244-276/Domain: intracellular #status predicted <IN3>  
F:277-300/Domain: transmembrane #status predicted <TM6>  
F:301-308/Domain: extracellular #status predicted <EX4>  
F:309-332/Domain: transmembrane #status predicted <TM7>  
F:333-373/Domain: intracellular #status predicted <IN4>  
F:20/196/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:319/Binding site: retinal (Lys) (covalent) #status predicted  
OOFF Length: 373 February 11, 2000 15:51 Type: P Check: 4358 ..  
  
1 MESFAVAQAQ LGPHFALSN GSVDKVTPD MAHLISPYNN QEPAPDPIMA  
51 KIIFAYMIMI GMISMGNGCV VIYIFATKS LRTPANILVI NLAIISFGIM  
101 ITNPMMGIN IVEETWLGP MMCDIYAGLG SAFGCSSIMS MCMSLDRYQ  
151 VIIVGMGNRP MTIPPLAKGI AVYMFSIW CLAPAGWR YYPEENLISC  
201 GIDLLEDWN PRYLIFYSI FYVYIPLELI CYSTWIILA VSARKARE  
251 OAKMANXS RSSEDKERSA EGKLAKVALV TITLMAMT PYLVINCGL  
301 FKREGULP LN TWIGACFAS AACNPIYVG ISHPKTRLA KEKCPCVF  
351 KVDDGKSSDA OSOTASEAE SKA  
11AA\_SEQUENCE 1.0  
  
Pl:155604 - platelet glycoprotein Ib beta chain precursor, endothelial aplice form - human  
N:Alternate names: membrane glycoprotein Ib beta chain (gpIb)  
C:Species: Homo sapiens (man)  
C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 22-Jun-1999  
C:Accession: U55604  
R:Kelly, M.D.; Essex, D.W.; Shapiro, S.S.; Meloni, F.J.; Druck, T.; Huebner, K.; Konte, B.A.  
J. Clin. Invest. 97, 2417-2424, 1994  
A>Title: Complementary DNA cloning of the alternatively expressed endothelial cell glycoprotein Ib beta (gpIb beta) and localization of the gpIb beta gene chromosome 22.  
A:Reference number: U55604; MUID:94259799  
A:Accession: U55604  
A>Status: translated from GB/EMBL/DDBT  
A:Molecule type: mRNA  
A:Residues: 1-411 <KEU>  
A:Cross-references: GB:L20860; NID:g438627; PIDN:AAA20398.1; PID:g517494  
C:Genetics:  
A:Gene: GDB:GPLIB  
A:Cross-references: GDB:I28731; OMIM:138720  
A:Map position: 22q11.21-22q11.23  
A:Introns: 209/1  
C:Complex: heterodimer with platelet glycoprotein Ib alpha chain (PIR:NBUHA)  
C:Superfamily: platelet glycoprotein Ib beta chain; leucine-rich alpha-2-glycoprotein repeat homology; proteoglycan amino-terminal homology; proteoglycan carboxyl-terminal homology  
C:Keywords: alternative splicing; blood coagulation; duplication; glycoprotein; heterodimer; phosphoprotein; platelet membrane; thrombin binding; transmembrane protein; von Willebrand factor binding  
F:224-249/Domain: proteoglycan amino-terminal homology <PAH>  
F:251-284/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRP>  
F:292-338/Domain: proteoglycan carboxyl-terminal homology <CBH>  
F:353-377/Domain: transmembrane #status predicted <TRM>  
F:271/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:396/Binding site: phosphate (Ser) (covalent) #status predicted  
155604 Length: 411 February 11, 2000 15:51 Type: P Check: 4750 ..  
  
1 MIPSRTMR FLPVNAASC PQDRMTLVN VAAGRVLR PLRAGSGSL  
51 SGIRPATVC YLPQRASA SCFLARPQH CGRCGRGG ALSLCSPAY  
101 ASRCVRRAA VESWPAPVL ESGRAPGSL GRPGRLGALV VWIQGETW  
151 RLKGDFQAC GVVERELAG YRDGHOGUD GAGPAWVLR DVAVPADRS  
201 AYCGASLAG RGALSLLL LAPSPHAG CPABSCAGT LVDCGRRGT  
251 WASLPAFVP DTTELVTGN NTALPEPLL DALPALPTAH LGANPWRCOC  
301 RVLPRLAWIA GRPERARYD LRCAVAPLAR GLLYLAED ELRAACAPCP  
351 LCWGALAQL ALTGILLHA LLTVLLCLR RLRLRARAR AAARLSLTDP  
401 LYAERAGIDE S  
11AA\_SEQUENCE 1.0  
Pl:ORBOB - oligopeptide transport system permease protein oppb - Salmonella typhimurium  
C:Species: Salmonella typhimurium  
C>Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 16-Jul-1999  
C:Accession: B29333  
R:Hiles, I.D.; Gallagher, M.P.; Jamieson, D.J.; Higgins, C.F.  
J. Mol. Biol. 195, 125-142, 1987  
A>Title: Molecular characterization of the oligopeptide permease of Salmonella typhimurium.  
A:Reference number: A29333; MUID:88011222  
A:Accession: B29333  
A:Molecule type: DNA  
A:Residues: 1-306 <HIU>  
A:Cross-references: GB:X05491; NID:g47801; PIDN:CAA29040.1; PID:g47803  
C:Comment: This is one of the highly hydrophobic membrane-associated components

of the periplasmic binding protein-dependent transport system for peptides that are two to five amino acids long. This transport system is of importance to the nutrition of the organism; it is also essential for recycling of cell wall peptides.

C:Genetics:  
A:Gene: oppB  
A:Map position: 34 min

C:Function:  
A:Description: probably responsible for mediating passage of peptides across the cytoplasmic membrane

C:Superfamily: oligopeptide permease protein oppB

C:Keywords: binding protein-dependent; transport system; oligopeptide transport;

transmembrane protein

F:32/Domain: transmembrane #status predicted <TM1>

F:105-121/Domain: transmembrane #status predicted <TM2>

F:140-156/Domain: transmembrane #status predicted <TM3>

F:173-189/Domain: transmembrane #status predicted <TM4>

F:228-244/Domain: transmembrane #status predicted <TM5>

F:273-289/Domain: transmembrane #status predicted <TM6>

OREBOB Length: 306 February 11, 2000 15:51 Type: P Check: 4920 ..

1 MKRFLIRCL EAIPTLFLIL TISFFMRLA PGSPTEGRA LPPEVLANIE

51 AKYHLNDPIM TOYFSYLKQL AHGDFGSPFK YKDYVNDLV AASEFVSACL

101 GAAMFLAVI IGVSAVIAA LKONTREWDYT VNGFAMTGVV IPSEFVAPLL

151 VNVFAITLQW LPGGWMNGA LKEMILPMVA LSLAYIASIA RITRGSMEIV

201 LHSFIFITAR AKGLPMRRII FRHALKPAAL PVLSTWGPAL VGIIIGSMVI

251 ETIYGLPGIG QLFVNGALNR DYSLVLSILI LVGALTILFN AIYDVLYAVI

301 DPKIRY

1 IAA\_SEQUENCE 1.0

P1:567566 - Probable membrane protein YDL033c - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: hypothetical protein D2761

C:Species: *Saccharomyces cerevisiae*

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: S67566

R:Fullin, L.; Søren, A.M.; Laamanen, P.

Submitted to the Protein Sequence Database, July 1996

A:Reference number: S67560

A:Accession: S67566

A:Molecule type: DNA

A:Residues: 1-417 <PAUV>

A:Cross-references: EMBL:274081; NID:g1431011; PTDN:CAA98591.1; PID:e252987;

PID:g1431012; MIPS:YDL033c

A:Experimental source: strain S288C

C:Genetics:

A:Map position: 4L

C:Superfamily: probable membrane protein YDL033c

C:Keywords: transmembrane protein

F:29-45/Domain: transmembrane #status predicted <TM>

S67566 Length: 417 February 11, 2000 15:51 Type: P Check: 8415 ..

1 MLARINLIG RRSASPYRPO RLPAKFDNVI VAMSSGVDS VAAALFACEF

51 PNTGVMQON WSESQSLDP GKEPCYERDM RDVNRVAKHL NIRDVKNVE

101 ODYWDVPEP MLRGYSEGST PNPDGCKNF VKFGKIREML DEKGTGMYM

151 LVYGHYARVM QEMNGKGLH LLRSIYRPKD QSYLLSQINS TVLSLLPLPI

201 GHTRKEPYRD LAKYAGLPTA EKPSQSGICE VNNSSQGRKR NFKIKYLPSS

251 PEDITVDPOQ SGAKTWGRH DGLMSYITGO KVQISNPQAD PRYQGTWYVS

301 EKLRDTNELI IVRGDNPAL YSDTWRIENF SSIGPREDTI NAFONTGALT

351 LQRRSLQVPV QIKSCKLNRS ADNUDDITIH ASKQRAITPG QSCCLYIDR  
401 VLGSGRPSHV NNNDTHA

1 IAA\_SEQUENCE 1.0

P1:S30896 - virulence-associated protein mbal - *Salmonella choleraesuis* plasmid

PKDSC50

C:Species: *Salmonella choleraesuis*

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: S30896; S20733

R:Matsu, H.; Abe, A.; Suzuki, S.; Kijima, M.; Tamura, Y.; Nakamura, M.;

Kawahara, K.; Danbara, H.

Mol. Gen. Genet. 236, 219-226, 1993

A:Title: Molecular mechanism of the regulation of expression of plasmid-encoded

mouse bacteremia (mba) genes in *Salmonella* serovar *Choleraesuis*.

A:Reference number: S30896; MUID:93173095

A:Accession: S30896

A:Molecule type: DNA

A:Residues: 1-297 <MAT>

A:Cross-references: EMBL:X54148; NID:g48765; PTDN:CAA38087.1; PID:g48766

C:Genetics:

A:Gene: mbal

C:Superfamily: virulence-associated protein spvr

C:Keywords: DNA binding; transcription regulation

S30896 Length: 297 February 11, 2000 15:51 Type: P Check: 7059 ..

1 MDPLINKLK IFTLLETGS FSATSVLYI TRTPLSRVS DLEELKORL

51 FIRNGLILP TEPAOTIYRK VKSHYIFLHA LEOELGPPGK TKOLEIIFDE

101 IYPSLNLKI ISALITSGOK TNMRAVNS QIIEELCOTN NCIVASRNY

151 FHRESIVCRT SVEGVMLEPI PKKFFLGKP DINRLAGTIV LFHEGANNFN

201 LDITVHFEEQ TIGITNPAPS EDNVLESSL YRLQGLANL LIPVAVCRAL

251 GLSTDHALHI KGVALTCTSLY YPKRKRETPD YRKAKILQOQ ELKOSTF

1 IAA\_SEQUENCE 1.0

P1:S06670 - virulence-associated protein, mKac - *Salmonella typhimurium* plasmid

N:Alternate names: 33K virulence protein spvr

C:Species: *Salmonella typhimurium*

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: S06670; A41318; S06089; S23714; S15213; C54540

R:Alvira, S.; Phen, M.

FEBS Lett. 257, 274-278, 1989

A:Title: Molecular organization of genes constituting the virulence determinant

on the *Salmonella typhimurium* 96 kilobase pair plasmid.

A:Reference number: S06670; MUID:90060335

A:Accession: S06670

A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-297 <TAIV>

A:Experimental source: virulence plasmid

R:Caldwell, A.L.; Guillig, P.A.

J. Bacteriol. 173, 7176-7185, 1991

A:Title: The *Salmonella typhimurium* virulence plasmid encodes a positive

regulator of a plasmid-encoded virulence gene.

A:Reference number: A41318; MUID:92041614

A:Accession: A41318

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-295 <CAL>

A:Cross-references: GB:M74110

A:Experimental source: strain SR-11, virulence plasmid

R:Pullinger, G.D.; Baird, G.D.; Williamson, C.M.; Lax, A.J.

Nucleic Acids Res. 17, 7983, 1989

A:Title: Nucleotide sequence of a plasmid gene involved in the virulence of

*Salmonella*.

A:Reference number: S06089; MUID:90016881

A:Accession: S06089

A:Molecule type: DNA  
 A:Residues: 1-297 <POL>  
 A:Cross-references: EMBL:X16111; NID:947859; PIDN:CAA34244.1; PID:947860  
 A:Experimental source: isolate 1275 wild-type, virulence plasmid  
 R:Note: F.; Pisano, M.R.; Nicoli, J.; Popoff, M.Y.  
 Res. Microbiol. 140, 627-630, 1989  
 A:Title: A plasmid-borne virulence region (2.8 kb) from *Salmonella typhimurium* contains two open reading frames.  
 A:Reference number: S23714; MUID:90176091  
 A:Accession: S23714  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-297 <NOR>  
 A:Cross-references: EMBL:X57092; NID:947788; PIDN:CAA0371.1; PID:947789  
 A:Experimental source: strain C5, virulence plasmid  
 R:Krusse, M.; Koudler, C.; Fierer, J.; Harwood, J.; Guiney, D.  
 Mol. Microbiol. 5, 307-316, 1991  
 A:Title: Molecular analysis of the virulence locus of the *Salmonella dublin* plasmid pSD2.  
 A:Reference number: S15213; MUID:91251759  
 A:Accession: S15213  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-40, 'G', 42-103, 'E', 105-297 <KRA>  
 A:Cross-references: EMBL:X56727; NID:947836; PIDN:CAA0047.1; PID:947837  
 R:Taara, S.; Baumann, M.; Rikonen, P.; Sukupolvi, S.; Rhen, M.  
 FEMS Microbiol. Lett. 77, 319-324, 1991  
 A:Title: Amino-terminal sequence analysis of four plasmid-encoded virulence-associated proteins of *Salmonella typhimurium*.  
 A:Reference number: A54540  
 A:Accession: C54540  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 1-10 <TA2>  
 C:Genetics:  
 A:Gene: mkac; mfc; spvr  
 A:Genome: plasmid  
 C:Superfamily: virulence-associated protein spvr  
 C:KeyWords: DNA binding; transcription regulation  
 S06670 Length: 297 February 11, 2000 15:51 Type: P Check: 7032 ..

1 MDLIRKRLK IFTLMETGS FSIASVLYI TRPLSRVIS DLREKQRL  
 51 FIRKNGTLP TEFQTIYRK VKSHYFLHA LEOEIGPTGK TKOLEIFDE  
 101 IYPSLKNLI ISALTISGQK TNIMGRAVNS QIEELCOTN NCIVISARNY  
 151 FHRESLVCRF SVEGVWLFI PKKFLGKRP DINRLAGTPV LFHEGAKNFN  
 201 LDTIHFPEQ TLGTTNPAFS FQNVDFSSL YRLQGGTAML LIVRYCRAL  
 251 GLSTHAIHI KVALCTSLY YPTKKRTPD YRKAIKLIQO ELKQSF

!!AA:SEQUENCE 1.0  
 P:IMCNA: colicin N immunity protein - *Escherichia coli* plasmid pCHAP4  
 C:Species: *Escherichia coli*  
 C:Date: 31-Mar-1991 #sequence-revision 31-Mar-1991 #text-change 16-Jul-1999  
 C:Accession: S01761  
 R:Pugsley, A.P.  
 Mol. Gen. Genet. 211, 335-341, 1988  
 A:Title: The immunity and lysis genes of *Colin* plasmid pCHAP4.  
 A:Reference number: S01760; MUID:88174431  
 A:Accession: S01761  
 A:Molecule type: DNA  
 A:Residues: 1-131 <PUG>  
 A:Cross-references: EMBL:X06933; NID:941117; PIDN:CAA30020.1; PID:941119  
 A:Note: the authors translated the codon ACC for residue 42 as Ile and GTT for residue 104 as Phe  
 C:Genetics:  
 A:Gene: cni  
 C:Superfamily: colicin N immunity protein

C:Keywords: transmembrane protein  
 F:66-84/Domain: transmembrane #status predicted <TM1>  
 IMCNA Length: 131 February 11, 2000 15:51 Type: P Check: 6528 ..

1 MHNTLEKII AYSLPQFHS LNNPLSEAF NLVYHTPLA ATSLFTTRK  
 51 ELKPKRSP LRAALKITLP TIIYISMIVC FILTDELTL SSXTEVLIYK  
 101 KRSEVFELY NTIYWDIYH IFVLVPRN I

!!AA:SEQUENCE 1.0  
 P:ZMECD - proteinase umud (EC 3.4.21.-) - *Escherichia coli*  
 N:Alternate names: umud; protease  
 C:Species: *Escherichia coli*  
 C:Date: 04-Dec-1986 #sequence-revision 04-Dec-1986 #text-change 16-Jul-1999  
 C:Accession: A03551; R23157; D64864  
 R:Kitagawa, Y.; Akaboshi, E.; Shinagawa, H.; Horii, T.; Ogawa, H.; Kato, T.  
 Proc. Natl. Acad. Sci. U.S.A. 82, 4356-4360, 1985  
 A:Title: Structural analysis of the umu operon required for inducible mutagenesis in *Escherichia coli*.  
 A:Reference number: A03550; MUID:85242679  
 A:Accession: A03551  
 A:Molecule type: DNA  
 A:Residues: 1-139 <KIT>  
 A:Cross-references: GB:M1387; NID:9148124; PIDN:AAA24728.1; PID:9148125  
 R:Perry, K.L.; Ellledge, S.J.; Mitchell, B.B.; Marsh, L.; Walker, G.C.  
 Proc. Natl. Acad. Sci. U.S.A. 82, 4331-4335, 1985  
 A:Title: umudC and umudB operators whose products are required for UV light- and chemical-induced mutagenesis: umud, umuA and LexA proteins share homology.  
 A:Reference number: A23157; MUID:85242678  
 A:Accession: A23157  
 A:Molecule type: DNA  
 A:Residues: 1-139 <PRY>  
 A:Cross-references: GB:M1387; NID:9148127; PIDN:AAA98073.1; PID:9148128  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Klapetich, H.A.; Goeden, M.A.; Rose, D.O.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of *Escherichia coli* K-12.  
 A:Reference number: A64720; MUID:97426617  
 A:Accession: D64864  
 A:Molecule type: DNA  
 A:Residues: 1-139 <BAT>  
 A:Cross-references: GB:AE000216; GB:U00096; NID:91787417; PIDN:AACT4267.1; PID:91787431; UMDP:D1183  
 A:Experimental source: strain K-12, substrain MG1655  
 C:Genetics:  
 A:Gene: umud  
 A:Map position: 26 min  
 C:Complex: umud/umud homodimer; umud'/umud' homodimer; umud/umud' heterodimer  
 C:Function:  
 A:Description: essential for induced and SOS mutagenesis in *E. coli*; controlled by *RecA* and *LexA* gene products; self cleavage of umud to umud' (mutagenetically active form); acts in concert with umudC, *RecA* and DNA polymerase III to facilitate the process of translesion synthesis, which results in the introduction of mutations  
 A:Note: one of two proteins encoded by umu operon; intracellular levels of umud and umud' proteins are kept to a minimum by *lon* serine proteinase; umud' is removed from the cell by the *ClpXP* serine protease, but only when it is in a heterodimer complex with umud  
 C:Superfamily: LexA repressor  
 C:Keywords: DNA repair; heterodimer; homodimer; hydrolase; induced mutagenesis; serine proteinase; SOS mutagenesis  
 F:1-139/Product: umud protein #status predicted <MUT1>  
 F:1-24/Domain: propeptide #link MAT2 #status predicted <PRO>  
 F:25-139/Product: umud' protein #status predicted <MAT2>

ZMECD Length: 139 February 11, 2000 15:51 Type: P Check: 8403 ..

1 MLIKPADLR EIVTFPLFS LVQCGFSPA ADYVEQRIDL NQLLIHPRA

51 TYFVKSAGDS MIDGDISDGD LLVDSAITA SHGDIVIAAV DGFTYKILQ  
 101 LRPTVOLIPW NSAYSPITIS SEPTLDVFCV VHVYKAKR

11AA\_SEQUENCE 1.0  
 PI:H64563 - flagellar motor switch protein - Helicobacter pylori (strain 26695)  
 C:Species: Helicobacter pylori  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: H64563  
 R:Tomb, J.F.; White, O.; Kellavag, A.R.; Clayton, R.A.; Sutton, G.G.;  
 Fleischmann, R.D.; Ketchum, K.A.; Klenk, H.P.; Dougherty, B.A.;  
 Nelson, K.; Quackenbush, J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus,  
 B.; Richardson, D.; Dodson, R.; Khaliq, H.G.; Glodek, A.; McKenney, K.;  
 Fitzgerald, L.M.; Lee, N.; Adams, M.D.; Hickey, E.K.; Berg, D.E.; Gocayne,  
 J.D.; Uitterlbeck, T.R.; Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Meldman,  
 J.M.; Fujii, C.; Bowman, C.; Watthey, L.  
 Nature 388: 539-547, 1997  
 A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.;  
 Fraser, C.M.; Venter, J.C.  
 A:Title: The complete genome sequence of the gastric pathogen Helicobacter  
 pylori.  
 A:Accession: H64563  
 A:Reference number: A64520; MUID:97394467  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1343 <DOM>  
 A:Cross-references: GB:AE000552; GB:AE000511; NID:92313451; PIDN:AD07420.1;  
 PID:92313453; TIGR:HP0352  
 C:Superfamily: flagellar switch protein filig  
 H64563 length: 343 February 11, 2000 15:51 Type: P Check: 9038

1 MATKLPKOR AOLDELMSSE KIAILLIQVG EDTGELIRH LIDISTEIS  
 51 KOIVOLNGTD KOIGAAVLEE FFAIQSNQY INTGLEVAR ELLFRILGSE  
 101 EAKVMDKLT KSLQTOKNFA YLGKIKPOL ADPIINEHQ TIALILAME  
 151 APNAETLSY FPEMKAEIS IRMANLEIS POYVKRSTV LNNLESITS  
 201 KYIEVGJRA VAEIFNRLQ KSATTLARI ESYDNLAGA IKEMTFED  
 251 IVKIDNFAIR ELKVADKRD LSLAKTSTK DLTDFLNMM SSRAGEFVE  
 301 EMQYGAVKI KDYVDAQRI IEIVSLOEK GVITGEED VIE

11AA\_SEQUENCE 1.0  
 PI:VEECU - unci protein - Escherichia coli  
 N:Alternate names: ATP synthase subunit atpI  
 C:Species: Escherichia coli  
 C>Date: 02-Apr-1982 #sequence\_revision 15-Oct-1982 #text\_change 16-Jul-1999  
 C:Accession: D30389; A23223; A34908; D65177; A04433  
 R:Walker, J.E.; Gay, N.V.; Saraste, M.; Eberle, A.N.  
 Biochem. J. 224: 799-815, 1984  
 A:Title: DNA sequence around the Escherichia coli unc operon. Completion of the  
 sequence of a 17 kilobase segment containing asna, orfC, unc, gims and phos.  
 A:Reference number: A30389; MUID:85121806  
 A:Accession: D30389  
 A:Molecule type: DNA  
 A:Residues: 1130 <WAL>  
 A:Cross-references: GB:X01631; NID:943256; PIDN:CAA25775.1; PID:9581248  
 R:Kanazawa, H.; Kiyasu, T.; Noumi, T.; Futai, M.  
 J. Bacteriol. 158: 300-306, 1984  
 A:Title: Overproduction of subunit a of the F0 component of  
 proton-translocation ATPase inhibits growth of Escherichia coli cells.  
 A:Reference number: A23223; MUID:84185438  
 A:Accession: A23223  
 A:Molecule type: DNA  
 A:Residues: 1130 <RAN>  
 R:Schnepp, B.; Deckers-Hebestreit, G.; Alendor, K.  
 J. Biol. Chem. 265: 389-395, 1990  
 A:Title: Overproduction and purification of the unci gene product of the ATP  
 synthase of Escherichia coli.

A:Reference number: A34908; MUID:90094427  
 A:Accession: A34908  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 712 <SCH>  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;  
 Riley, M.; Colladon-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,  
 J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277: 1453-1462, 1997  
 A:Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:97426617  
 A:Accession: D65177  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1130 <BLAT>  
 A:Cross-references: GB:AE000450; GB:U00096; NID:91790166; PIDN:AAC76762.1;  
 PID:91790177; UWGP:b3739  
 A:Experimental source: strain K-12, substrain MG1655  
 R:Kanazawa, H.; Mabuchi, K.; Futai, M.  
 Biochem. Biophys. Res. Commun. 107: 568-575, 1982  
 A:Title: Nucleotide sequence of the promoter region of the gene cluster for  
 proton-translocating ATPase from Escherichia coli and identification of the  
 active promoter.  
 A:Reference number: I52197; MUID:83022404  
 A:Accession: I52197  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 747, 32-84 <RES>  
 A:Cross-references: EMBL:V00313; NID:942286; PIDN:CAA23599.1; PID:9581172  
 C:Comment: The unci gene is the first gene of the unc (or atp) operon. A  
 possible function for this protein is to guide the assembly of the membrane  
 sector of the proton transporting ATP synthase (EC 3.6.1.34) enzyme complex.  
 C:Genetics:  
 A:Gene: atpI; unci  
 A:Map position: 84 min  
 A:Start codon: GAG  
 C:Superfamily: unci protein  
 C:Keywords: transmembrane protein  
 VEECU length: 130 February 11, 2000 15:51 Type: P Check: 4271

1 MKNVSVSLV SRNVARKLL VOLLVIASG LFLSLDPW GVSASGLA  
 51 VFLPVLFLMI FAWRHOATP AKGRVATFA FGEAFKVLAM LVLLVALAV  
 101 IKAFLPLIV TWVLVVOI LAPAVINNG

11AA\_SEQUENCE 1.0  
 PI:8AG55 - virB8 protein precursor - Agrobacterium tumefaciens plasmid pTi15955  
 C:Species: Agrobacterium tumefaciens  
 C>Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 16-Jul-1999  
 C:Accession: S00784  
 R:Thompson, D.V.; Melchers, L.S.; Idler, K.B.; Schilperoord, R.A.; Hooykaas,  
 Nucleic Acids Res. 16: 4621-4636, 1988  
 A:Title: Analysis of the complete nucleotide sequence of the Agrobacterium  
 tumefaciens virB operon  
 A:Reference number: S00777; MUID:88247765  
 A:Accession: S00784  
 A:Molecule type: DNA  
 A:Residues: 1257 <TBO>  
 A:Cross-references: EMBL:X06826; NID:939195; PIDN:CAA23978.1; PID:939210  
 C:Genetics:  
 A:Gene: plasmid  
 C:Superfamily: tumor-inducing plasmid pTiC58 virB8 protein  
 F:1-36/Domain: signal sequence #status predicted <SIG>  
 F:137-257/Product: virB8 protein #status predicted <WAT>  
 8AG55 length: 257 February 11, 2000 15:51 Type: P Check: 173

1 MMGDSLRQ IFSAIRVDA MTGPYAMLV ARSLERHK EYEAFOIARA  
 51 KSARLSKII AAVAAIILG NVAQFAIAT MPPLSLRVPV YLMIRPDGTV

101 DSEVSISRLP ATOEAVVNA SIMEYVRLR SYDADTAQYA YDLVSNFSAP  
 151 TVRQDYCOOF NYPNPSSPOV ILGKRGREV EHIASNDVTP STQOIRYKRT  
 201 LVDDKMPV STWTATVRE KVTSLPGRLR LTNPAGLVVT SYQSEDTVS  
 251 NVGQAGP

## !!AA\_SEQUENCE 1.0

P1:BAG58 - virB8 protein - Agrobacterium tumefaciens plasmid pTiA6

C:Species: Agrobacterium tumefaciens

C>Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 03-Feb-1994

C:Accession: J28621; J27127

R:Ward, J.E.; Akiyoshi, D.E.; Regier, D.; Datta, A.; Gordon, M.P.; Nester, E.W.

J. Biol. Chem. 265, 5804-5814, 1988

A:Title: Characterization of the virB operon from an Agrobacterium tumefaciens

11 plasmid.

A:Reference number: A28621; MUID:88186901

A:Accession: J28621

A:Molecule type: DNA

A:Residues: 1-230 <MAR>

A:Cross-references: GB:J03216

A>Note: this sequence was designated ORF 9 in this reference

R:Ward, J.E.; Akiyoshi, D.E.; Regier, D.; Datta, A.; Gordon, M.P.; Nester, E.W.

J. Biol. Chem. 265, 4768, 1990

A:Reference number: A35737; MUID:90170994

A:Contents: annotation; erratum

C:Genetics:

A:Genome: plasmid

C:Superfamily: tumor-inducing plasmid pTiC58 virB8 protein

BAG58 Length: 230 February 11, 2000 15:51 Type: P Check: 2646 ..

1 MIVANESLAE HYKEVEAFQT ARKSKRRIS KIIAVAIA ILGNVACAPA

51 IATWPLSLRL VPYVLMIRAD GTVDSEVSIS RLPAQGEAV VNASLMEIYR

101 LRESDADTA QYAYDLVSNF SAEFTVADYQ QFENTPNPSS PQVILGKRG

151 VEVEHIASND VTPSTQOIRY KRLVVYDGM PVSTWTATV RYKVTSLPG

201 RLRLTPAGL VVTSYQTSQD TVSNVQAGAP

!!AA\_SEQUENCE 1.0

P1:BAG58 - virB8 protein - Agrobacterium tumefaciens plasmid pTiC58

C:Species: Agrobacterium tumefaciens

C>Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 16-Jul-1999

C:Accession: S12348; S11833; S10523

R:Shirasu, K.; Morel, P.; Kado, C.I.

Mol. Microbiol. 4, 1153-1163, 1990

A:Title: Characterization of the virB operon of an Agrobacterium

plasmid: nucleotide sequence and protein analysis.

A:Reference number: S12348; MUID:91041724

A:Accession: S12348

A:Molecule type: DNA

A:Residues: 1-237 <SHI>

A:Cross-references: EMBL:J03320; NID:9154781; PID:AAA91598.1; PID:9154790

R:Rogovsky, P.M.; Powell, B.S.; Shirasu, K.; Lin, T.S.; Morel, P.; Zyprian,

E.M.; Steck, T.R.; Kado, C.I.

Plasmid 23, 85-106, 1990

A:Title: Molecular characterization of the virB8 protein of Agrobacterium

tumefaciens: complete nucleotide sequence and gene organization of the

28.63-kbp region cloned as a single unit.

A:Reference number: S11825; MUID:90301800

A:Accession: S11833

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-237 <ROG>

A:Cross-references: EMBL:J03320; NID:9154781; PID:AAA91598.1; PID:9154790

R:Kulda, G.A.; de Vos, G.; Owen, J.; McCaffrey, G.; Zambryski, P.

Mol. Gen. Genet. 221, 256-266, 1990

A:Title: The virB operon of Agrobacterium tumefaciens pTiC58 encodes 11 open

## reading frames.

A:Reference number: S10516; MUID:90318324

A:Accession: S10523

A:Molecule type: DNA

A:Residues: 1-20,23-127, 'SA', 130-237 <KUL>

A:Cross-references: EMBL:X53264; NID:939152; PID:CAA37361.1; PID:939160

C:Genetics:

A:Gene: virB8

A:Genome: plasmid

C:Superfamily: tumor-inducing plasmid pTiC58 virB8 protein

BAG58 Length: 237 February 11, 2000 15:51 Type: P Check: 6263 ..

1 MKGSEYALLV ARETLAEHYK EYEAFTOTARA KSARLSKYI AAVATIAVLG

51 NVAQAFITAT MVLIRLVPV YLMIRPDGV DSEVSISRLP ATOEAVVNA

101 SIMEYVRLR SYDADTAQYA YDLVSNFTRP MYRONYCOOF NYPNPSSPOV

151 ILGKGRLEV EHIASNDVTP GVOQIRYKRT LIVDKMPMA STWTATVRE

201 KVTSLPGRLR LTNPGGLVVT SYQSEDTVS NAGHSEP

!!AA\_SEQUENCE 1.0

P1:W2LJ31 - E2 protein - human papillomavirus type 31

C:Species: human papillomavirus type 31

A:Note: host Homo sapiens (man)

C>Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 16-Jul-1999

C:Accession: D32444

R:Goldsbrough, M.D.; Disilvestre, D.; Temple, G.F.; Lorincz, A.T.

Virology 171, 306-311, 1989

A:Title: Nucleotide sequence of human papillomavirus type 31: a cervical

neoplasia-associated virus.

A:Reference number: A94398; MUID:89299478

A:Accession: D32444

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-372 <GOL>

A:Cross-references: GB:U04353; NID:9333048; PID:AAA6995.1; PID:9459919

C:Superfamily: Papillomavirus E2 protein

C:Keywords: DNA binding; early protein; transcription regulation

W2LJ31 Length: 372 February 11, 2000 15:51 Type: P Check: 9656 ..

1 METLSQRLNV CODKLEHYE NDSKRLCDHI DYMKHRLLEC VLMYKAREMG

51 IHSIHQVVP ALSYSKAKAL QALEQOMLE TLNNTYKNE DMTMOQTSLE

101 LYLTPATGCL KKHGYVEVQ FGDVHNTWMH YTNKRFIYLC IDQOCTVVG

151 QVNCAGIYV HEGHIITYFN FTEBAKKYGT GKKEVHAGG QVAFPESEVF

201 SDELSFPGI VTKLPTANNT TTSNSKTCAL GTSCGVRRAT TSKRRTPE

251 EHRNTHPRNK LRGDSVDSV NCGVISAAC TNOTRAVSCP ATPPIHLKG

301 DANILKCLRY RLSKRYQLYE QVSTHWHWC TDGKHNAIV TLTYYISTOR

351 DFLNTVXIP NTVSYSTGYM TI

!!AA\_SEQUENCE 1.0

P1:W2LJ31 - E2 protein - human papillomavirus type 51

C:Species: human papillomavirus type 51

A:Note: host Homo sapiens (man)

C>Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 16-Feb-1997

C:Accession: B40415

R:lungu, O.; Crum, C.P.; Silverstein, S.J.

J. Virol. 65, 4216-4225, 1991

A:Title: Biologic properties and nucleotide sequence analysis of human

papillomavirus type 51

A:Reference number: A40415; MUID:91303675

A:Accession: B40415

A:Status: translation not shown



Mon Feb 14 08:07:32 2000

pir.cat

Page 61

```
A:Molecule type: DNA
A:Residues: 1-358 <LUN>
A:Cross-references: GB:M62877
C:Superfamily: Papillomavirus E2 protein
C:Keywords: DNA binding; early protein; transcription regulation
W2WL51 Length: 358 February 11, 2000 15:51 Type: P Check: 1816

1 METLCRLN COEKLIDCE LDSKLVDOI NWTLIRREA AMFYAERN
51 LRTINQVAV ATTYSKQAC QAIMHVALO SLNKSDYNNK PMTRETCE
101 LMCVAKQCF KKGITVIV FDKGNKNAMD YSMKFIYV DNDKRWKNG
151 NVDYGIYV VNSKREYVQ FKDEAKIYGA QOMEVYMTGT VITCEPVSS
201 TCSDALSTT TVEQLSTPT TNPITTCVGA KEAQOQKRR QRLTEPDSST
251 ISPLSYDNTN NQIHGSGST NTGHSOATO TAFIYHLKGD TNCICKFRYR
301 FTKHKLGYKN VSTWMTSN TKGTIVTVF DSAHQRETFI KTIKVPSTV
351 LSLGIMTL

!!A_SEQUENCE 1.0
P1:W2WL18 - E2 protein - human papillomavirus type 18
C:Species: human papillomavirus type 18
C>Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 16-Jul-1999
R:Coile, S.T.; Danos, O.
J. Mol. Biol. 193, 599-608, 1987
A:Title: Nucleotide sequence and comparative analysis of the human
papillomavirus type 18 genome. Phylogeny of papillomaviruses and repeated
structure of the E6 and E7 gene products.
A:Reference number: A92937; MUID:87283862
A:Accession: D26251
A:Molecule type: DNA
A:Residues: 1-365 <COL>
A:Cross-references: GB:X05015; NID:960975; PIDN:CAA28667.1; PID:960979
R:Meisner, J.
Nucleic Acids Res. 21, 1041, 1993
A:Title: Tag1 is a single cut enzyme for HPV-18.
A:Reference number: S35634; MUID:93197132
A:Accession: S35634
A:Molecule type: DNA
A:Residues: 84-89, 'A', 91-93 <MEI>
A:Cross-references: EMBL:L07917
C:Superfamily: Papillomavirus E2 protein
C:Keywords: DNA binding; early protein; transcription regulation
W2WL18 Length: 365 February 11, 2000 15:51 Type: P Check: 5567

1 MOPKRETLSE RLSCVODKII DHYENDSKDI DSOIQWQII RMENAFPA
51 REHGIOFLNH QVPAVINISK SKAKATEIQ MALOGIACSR YKTEDWTLQD
101 TCEELMWTPE THCFKKGQT VOYFEDNKN DCMITYAMDS VYMTDAGTW
151 DKATVCVSR GLYVKEGN TFFTEFKSEC EKYGNNGTGE VHFNNVYDC
201 NDSMCSSTD TVSATQVLQO LQHTPSYSS TVSVGAKTY GQTSAAIRPG
251 HGLAEKQRC GFVNPLGAA TPTGNKRRK LCSGNTTPII HLKDRNSLK
301 CLNRRLRKS DHYRDISSTW HMTGAGNEKT GILTVYHSE TORFKFLNTV
351 AIRDSVOILV GYMTM

!!A_SEQUENCE 1.0
P1:SL5617 - E2 protein - human papillomavirus type 2a
C:Species: human papillomavirus type 2a
A:Note: seq Homo sapiens (man)
C>Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 16-Feb-1997

C:Accession: SL5617
R:Hirsch-Behnam, A.; Delius, H.; de Villiers, E.M.
Virus Res. 18, 81-98, 1990
A:Title: A comparative sequence analysis of two human papillomavirus (HPV)
types 2a and 57
A:Reference number: SL5614; MUID:91188699
A:Accession: SL5617
A:Molecule type: DNA
A:Residues: 1-391 <HR>
A:Cross-references: EMBL:X55964
C:Superfamily: Papillomavirus E2 protein
C:Keywords: DNA binding; early protein; transcription regulation
SL5617 Length: 391 February 11, 2000 15:51 Type: P Check: 734

1 METLANRDA COETLLELYE KDSNLEDOI KHWAOVLEN VMLFKARECG
51 MTRVGTAVP ALTVSKAKAC QAEVQALQ TLMQSYSTE AMTLDTCLE
101 MMDAPPKCW KKGQSVLYK FDGSSDRDMI YSMGFIVQ DTIDSMHRY
151 PGVDELGLY YVHDGVNRY VDPETESLY GVTGWEVAV AGTVIHTSA
201 SVSSTQASAS DDEPLSPIRT AVSPVPAPVA ASAEISGAGR AAPPTQALCS
251 AQAPTSPAK RQRYVGOQH PRPSTRTYG EGEVECYNKR SIDSNTDP
301 RWGHDITDV PYIHLRGDAN CLKCFRYRQ KKKDVLXARY SSTWMAAGN
351 GDKTAEFTLV YTSVEQTEF LTRYSIPKGL IALPGYMSAF V

!!A_SEQUENCE 1.0
P1:QOBN5 - Integral membrane protein - saimiriine herpesvirus 1 (strain 11)
C:Species: saimiriine herpesvirus 1
A:Note: host Saimiri sciureus (common squirrel monkey)
C>Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jul-1999
C:Accession: A36810
R:Albrecht, J.
submitted to the EMBL Data Library, January 1992
A:Description: Primary structure of the herpesvirus saimiri genome.
A:Reference number: A36806
A:Accession: A36810
A:Molecule type: DNA
A:Residues: 1-366 <ALB>
A:Cross-references: GB:X64346; NID:960320; PIDN:CAA45662.1; PID:960360
R:Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.; Biesinger, B.;
Newman, C.; Wiltmann, S.; Craxton, M.A.; Coleman, H.; Fleckenstein, B.; Honess,
R.W.
J. Virol. 66, 5047-5058, 1992
A:Title: Primary structure of the herpesvirus saimiri genome.
A:Reference number: A37309; MUID:92333688
A:Contents: annotation; possible protein-coding frames
A:Note: neither amino acid nor nucleotide sequence is given
C:Genetics:
A:Gene: 39
C:Superfamily: cytomegalovirus UL100 protein
C:Keywords: transmembrane protein
E:18-34/Domain: transmembrane #status predicted <TM>
E:87-103/Domain: transmembrane #status predicted <TM>
E:152-168/Domain: transmembrane #status predicted <TM>
E:214-230/Domain: transmembrane #status predicted <TM>
E:236-252/Domain: transmembrane #status predicted <TM>
E:274-290/Domain: transmembrane #status predicted <TM>
E:303-321/Domain: transmembrane #status predicted <TM>
QOBN5 Length: 366 February 11, 2000 15:51 Type: P Check: 7071

1 MMKASRSDTF MTRTWIQLV LEVIMTMSA IIPIASVSG LGFCYFPNL
51 VDYLLWTL RNAAKHLPT LLEAPELFV YITWSVLVDL ASAIYYVGA
101 LAILQARKTH LTSMTLQTM INVGSHTML FGIARMTML QLFTHVLSYK
```

```
151  HVMALAEIYF LHFCLSYMHT LSLVSRNSPK WSVLLMEQHI PKOSLSTILL
201  DYKRPICVNM YLSLLALEML VFSLGFMAI GNSFIIVSD TVLASINLXF
251  VLTETWYMMT EMFLQDYIYL QFGFYLGVS GSLILLFVL RYBAVVSAN
301  LKHTAVNIA MIPACVIAM MFLRFYSQO VKRPENSTYP LPRFKRRQ
351  KQOQILME TSDEEL

11AA_SEQUENCE 1.0
P1:WMNZ22 envelope-associated 22k protein - human respiratory syncytial virus
C:Species: human respiratory syncytial virus
C:Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 16-Jul-1999
C:Accession: B93010; A93009; A04034
R:Elango, N.; Satake, M.; Venkatesan, S.
J. Virol. 55, 101-110, 1985
A:Title: mRNA sequence of three respiratory syncytial virus genes encoding two
A:Functional proteins and a 22k structural protein.
A:Reference number: A93010; MUID:85237684
A:Accession: B93010
A:Molecule type: genomic RNA
A:Residues: 1-194 <ELAV>
A:Cross-references: GB:M11486; NID:933925; PID:AB59860.1; PID:9333934
R:Collins, P.L.; Wertz, G.W.
J. Virol. 54, 65-71, 1985
A:Title: The envelope-associated 22k protein of human respiratory syncytial
A:Virus: nucleotide sequence of the mRNA and a related polytranscript.
A:Reference number: A93009; MUID:85135082
A:Accession: A93009
A:Molecule type: genomic RNA
A:Residues: 1-194 <COL>
A:Cross-references: GB:M11486; GB:K01459; GB:K02719; GB:K03348; GB:K03349;
GB:M11217; GB:M1244; GB:M11487; GB:M11505; GB:M11514; GB:M11631; GB:M12966;
GB:X00001; GB:X02221; NID:9333925; PID:AB59860.1; PID:9333934
C:Genetics:
A:Gene: 22K
C:Superfamily: respiratory syncytial virus envelope-associated 22k protein
F:89,191/Binding site: carbohydrate (Asn) (covalent) #status predicted
WMNZ22 Length: 194 February 11, 2000 15:51 Type: P Check: 9828 ..

1 MSRRNPCKFE IRGHCLNGKR CHFSNHYEW PPHALLVRON FMNLRLKSM
51 DKSIDLSEI SGAELDRTE EYALGVYVL ESYIGSINNI TKOSACVAMS
101 KLTLENSDD IKRLDNEEL NSFIRVYNT VISIIESNRK NMQIHLK
151 RLPADVLKKT IKNTLDIHS ITINPKEST VSDINDAKN NDPT

11AA_SEQUENCE 1.0
P1:WMNZBA - matrix glycoprotein M2 - bovine respiratory syncytial virus (strain
A51908)
C:Species: bovine respiratory syncytial virus
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jul-1999
C:Accession: J01482
R:Zamorá, M.; Samal, S.K.
J. Gen. Virol. 73, 737-741, 1992
A:Title: Sequence analysis of M2 mRNA of bovine respiratory syncytial virus
A:obtained from an F-M2 distalronic mRNA suggests structural homology with that
A:of human respiratory syncytial virus.
A:Reference number: J01481; MUID:92185490
A:Accession: J01482
A:Molecule type: mRNA
A:Residues: 1-186 <ZAM>
A:Cross-references: GB:M82816; NID:9210823; PID:AAA42805.1; PID:9210825
C:Genetics:
A:Gene: M2
C:Superfamily: respiratory syncytial virus envelope-associated 22k protein
F:54,89/Binding site: carbohydrate (Asn) (covalent) #status predicted

WMNZBA Length: 186 February 11, 2000 15:51 Type: P Check: 6337 ..

1 MSRRNPCKFE IRGHCLNGKR CHFSNHYEW PPHALLVRON FMNLRLKSM
51 DRNDITLSEI SGAELDRTE EYALGVYVL ESYIGSINNI TKOSACVAMS
101 KLTLENSDD IKRLDNEEL NSFIRVYNT VISIIESNRK NMQIHLK
151 RLPADVLKKT IKNTLDIHS ITINPKEST VSDINDAKN NDPT

11AA_SEQUENCE 1.0
P1:WMNY22 12k protein - potato virus M (strain Russian)
C:Species: potato virus M
C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 23-Jul-1999
C:Accession: P00003; C54333; S21603
R:Rupakov, V.V.; Morozov, S.Y.; Kanyuka, K.V.; Zavrilev, S.K.
J. Gen. Virol. 70, 1861-1869, 1989
A:Title: Partial nucleotide sequence of potato virus M RNA shows similarities
A:to potexviruses in gene arrangement and the encoded amino acid sequences.
A:Reference number: A92800; MUID:89293091
A:Accession: P00003
A:Molecule type: mRNA
A:Residues: 1-109 <RUP>
A:Cross-references: GB:D14449; GB:ID00515; GB:X53062; NID:9222424;
PID:BA03341.1; PID:d1003851; PID:9222427
R:Zavrilev, S.K.; Kanyuka, K.V.; Levy, K.E.
J. Gen. Virol. 72, 9-14, 1991
A:Title: The genome organization of potato virus M RNA.
A:Reference number: A54333; MUID:91116326
A:Accession: C54333
A:Molecule type: genomic RNA
A:Residues: 1-109 <ZAV>
A:Cross-references: EXBL:X53062; NID:961291; PID:CA37334.1; PID:961294
C:Comment: The genome is a single-stranded, positive-sense RNA.
C:Superfamily: barley stripe mosaic virus 14k protein
C:Keywords: transmembrane protein
WMNY22 Length: 109 February 11, 2000 15:51 Type: P Check: 3415 ..

1 MFLTPPEPT KYVIALAGV SLALVWLLI RSTLPVVGDR DNNLPNGWY
51 RDGTRSVFYN SPGRINSTEIA RKAPLIGQP ALVVLVLLI WASHKIGRN
101 CRACAGSHT

11AA_SEQUENCE 1.0
P1:VCBWC - coat protein - tobacco streak virus (strain WC)
N:Alternate names: P4 protein
C:Species: tobacco streak virus TSV
A:Note: host Nicotiana glauca, N. Clevelandii (tobacco)
C:Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 23-Jul-1999
C:Accession: A04206
R:Corneilissen, B.J.C.; Janssen, H.; Zuidema, D.; Bol, J.F.
Nucleic Acids Res. 12, 2427-2437, 1984
A:Title: Complete nucleotide sequence of tobacco streak virus RNA 3.
A:Reference number: A93507; MUID:84169544
A:Accession: A04206
A:Molecule type: genomic RNA
A:Residues: 1-237 <COR>
A:Cross-references: GB:X00435; NID:962146; PID:CA25133.1; PID:962148
C:Genetics:
A:Map position: segment 3
C:Superfamily: tobacco streak virus coat protein
VCBWC Length: 237 February 11, 2000 15:51 Type: P Check: 3595 ..

1 NNTLIGGDH PSNAMSRRAN NRSNNSCPT CIDELDAMR NCPAHNTVNT
51 VSRQRRAA RAAAYRNANA RVPLPLPVVS VSRPQAKSL RLPNNQWVT
101 RASMSAKT VDTNDALPRK TIVEGPEIG AERKFRLLI GFVAVSDGT
```

```

151 GMDVGVTDV IDPPVVGRL GFKKNTYRSR DFDLGGKLLN QLDRAVVMC
201 LDERRRRAKR VOLAGYWINI SKPAPLMPPE DFLVNOD

!!AA_SEQUENCE 1.0
P1:VCVETC - coat protein - turnip crinkle virus
C:Species: turnip crinkle virus
A:Note: host Brassica rapa (turnip)
C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 23-Jul-1999
C:Accession: JAO111; S07285
R:Carlington, J.C.; Heaton, L.A.; Zuidema, D.; Hillman, B.I.; Morris, T.J.
Virology 170, 219-226, 1989
A:Title: The genome structure of turnip crinkle virus.
A:Reference number: A94394; MUID:8923179
A:Accession: JAO111
A:Molecule type: genomic RNA
A:Residues: 1-351 <CAR1>
R:Carlington, J.C.; Morris, T.J.; Stockley, P.G.; Harrison, S.C.
J.Mol. Biol. 194, 265-276, 1987
A:Title: Structure and assembly of turnip crinkle virus. IV. Analysis of the
coat protein gene and implications of the subunit primary structure.
A:Reference number: S07285; MUID:87283926
A:Accession: S07285
A:Molecule type: genomic RNA
A:Residues: 1-351 <CAR2>
A:Cross-references: EMBL:X05193; NID:962114; PIDN:CAA28623.1; PID:962115
A:Note: the authors translated the codon TTG for residue 346 as Trp
C:Comment: The genome is a single-stranded, positive-sense RNA.
C:Superfamily: carnation mottle virus coat protein
C:Keywords: coat protein
F:1-56/Domain: R <DOR>
F:62-246/Domain: S <DOS>
F:249-351/Domain: P <DOP>

VCVETC Length: 351 February 11, 2000 15:51 Type: P Check: 1818 ..

1 MENDPRVKRF ASDGQWAKR MOKKGMSTLT SROKOTARAR MGIKSPVAQ
51 PQVKTRLSA PVALAYREVS TOPRVSTARD GTRFSGSELI TLKKNITDE
101 PRTTAVANLP SEPTGFNQLI KEAQYKXFR FTSIRFRYSP MSPSTTGKRV
151 ALAFPRDAK PPPNDLASL NIEGCVSSVP WTGFTLVPT DSTDRFVADG
201 ISDPKLVDPG KLIMATYGQG ANDAQLGEV RVETVQLKN RFGSTSDAQI
251 GDFAGVKQCP RLYSMKTKG TAGMEHCHF LGTGNESLTL FTEKAPVSGL
301 ENADASDPSTV LGENAAGSVQ WAGVVAERK QGVKMTTEE QKRGKQLALR
351 I

!!AA_SEQUENCE 1.0
P1:VHPPF6 - major nucleocapsid protein - phage phi-6
N:Alternate names: P8 protein
C:Species: phage phi-6
A:Note: host Pseudomonas phage6
C:Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 23-Jul-1999
C:Accession: A23368
R:McGraw, T.; Mindich, L.; Frangione, B.
J.Virol. 58, 142-151, 1986
A:Title: Nucleotide sequence of the small double-stranded RNA segment of
bacteriophage phi6: novel mechanism of natural translational control.
A:Reference number: A93017; MUID:86144085
A:Accession: A23368
A:Molecule type: genomic RNA
A:Residues: 1-149 <MCG>
A:Cross-references: GB:M12921; NID:9215492; PIDN:AAA32358.1; PID:9215493
C:Genetics:
A:Gene: P8
A:Map position: segment S
C:Superfamily: phage phi-6 nucleocapsid protein
C:Keywords: nucleocapsid

VHPPF6 Length: 149 February 11, 2000 15:51 Type: P Check: 6252 ..

1 MLFPVARRA VPAIESAIA TPGLYSRIA AIGSKVSPSA ILAAVKNPV
51 VAGLTLLAQIG STGYDAYOOL LENHPEVAEM LKDSLKADE IQDPFGNIG
101 QYRELELVE DAARFVGMS NLIRLQALE LDIKIYGLKM QLNDMGTRS

!!AA_SEQUENCE 1.0
P1:ZGBP4 - gene G protein - phage G4
C:Species: phage G4
C:Date: 30-Nov-1979 #sequence_revision 30-Nov-1979 #text_change 23-Jul-1999
C:Accession: A04252
R:Godson, G.N.; Bartell, B.G.; Staden, R.; Fildes, J.C.
Nature 276, 236-247, 1978
A:Title: Nucleotide sequence of bacteriophage G4 DNA.
A:Reference number: A93200; MUID:79053264
A:Accession: A04252
A:Molecule type: DNA
A:Residues: 1-177 <GOD>
A:Cross-references: GB:J02454; GB:M10724; GB:M1404; GB:V00657; NID:9215415;
PIDN:AAA32324.1; PID:9215425
C:Comment: gene G protein is one of the structural components of the
bacteriophage capsid.
C:Superfamily: phage phi-X174 gene G protein

ZGBP4 Length: 177 February 11, 2000 15:51 Type: P Check: 4484 ..

1 MQKFTSKHN APINSTOLAA TTTPAVAPV LSPVNLRSRT ILINNTTAV
51 TTHSGICHHV RIDETNPTNH HALSTAGSL NVRADMTARA IFEVADGV
101 PTAVPALIDV YPIETENNGK AISEKDAVTI DSHPRVYGDN VYAGIMLSN
151 AMTASTISGV LSVQVNRRA TVLOPLK

!!AA_SEQUENCE 1.0
P1:PDYBC - REP2 protein - yeast (Saccharomyces cerevisiae) plasmid
N:Alternate names: hypothetical protein C
C:Species: Saccharomyces cerevisiae
C:Date: 31-Oct-1980 #sequence_revision 23-Oct-1981 #text_change 23-Jul-1999
C:Accession: A04504
R:Hartley, J.L.; Donelson, J.E.
Nature 286, 860-864, 1980
A:Title: Nucleotide sequence of the yeast plasmid.
A:Reference number: A93232; MUID:81012161
A:Accession: A04504
A:Molecule type: DNA
A:Residues: 1-296 <HAR>
A:Cross-references: GB:J01347; NID:9172190; PIDN:AAB59340.1; PID:9172191
A:Experimental source: strain A364A D5
C:Genetics:
A:Genome: plasmid
C:Superfamily: yeast REP2 protein
C:Keywords: plasmid partition

PDYBC Length: 296 February 11, 2000 15:51 Type: P Check: 6596 ..

1 MDDIETAKNL TVKARIVASV MDVCRFLFTM IAPVDIDIE SKRKSLELLF
51 PGYVIRPMES LTGRPPYGLD SSAEDSVSS DSADEVILPA AKWAKRRFDS
101 IGNGLMSOE ASQAIDML ONNKLNDNR QLYKSIAIII GRLEPKDKKR
151 ATEMLRKMD CTOLLVPAP TEEDVKKLV VVTQLTLVP PDRQALLID
201 LFIPESLKDI ENSFNEAAE NRILOKSEL EGRETVNHN TNEVPSRRT
251 RSRDINARGA YKLONTITEG PRAVPTKKRR VATRGRKRS RNTSRV

!!AA_SEQUENCE 1.0
P1:B64166 - hypothetical protein H11084 - Haemophilus influenzae (strain Rd

```

KW20)  
 C:Species: Haemophilus influenzae  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: B64166  
 R:Feilichmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.;  
 Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.; Dougherty, B.A.; Merrick, J.M.;  
 McEnney, K.; Sutton, G.; Fitzhugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;  
 Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.F.; Phillips, C.A.;  
 Spriggs, T.; Hedblom, E.; Cotton, M.D.; Uterback, T.R.; Hanna, M.C.;  
 Nguyen, P.T.; Saudek, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.;  
 Fuhmann, J.L.; Geoghagen, N.S.M.  
 Science 269, 496-512, 1995  
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.;  
 Venter, J.C.  
 A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae  
 Rd.  
 A:Reference number: A64000; MUID:95350630  
 A:Accession: B64166  
 A>Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-214 <TIGR>  
 A:Cross-references: GB:U32788; GB:I42023; NID:g1574629; PID:g1574638;  
 TIGR:H11084  
 C:Note: best homolog was a hypothetical protein from Escherichia coli  
 C:Superfamily: hypothetical protein H11084  
 B64166 Length: 214 February 11, 2000 15:51 Type: P Check: 1216 ..

1 MNLIQKWF TILFVLTAET LVTRALAEI SPYLVKQQA DKLFDIQAN  
 51 OSKIKDPNV LRTIVRNDL PYVLKEXAGS KVLGSYKST SAEQREKFX  
 101 TFEELIEQX AQLNYSNQ KIQEISEKL GDNFNINIV NIQANGVAP  
 151 ILIFYKRRK NKSSEKMYD MVGAGVSMLE DTKNVGLI NKQGDITIT  
 201 KMQSASQPI IFNQ  
 !:A SEQUENCE 1.0  
 P1:E70044 - conserved hypothetical protein yvob - Bacillus subtilis  
 C:Species: Bacillus subtilis  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: E70044  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo,  
 V.; Bartero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.;  
 Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;  
 Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.;  
 Conerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denicot, F.; Devine, K.M.;  
 Duesterhoeft, A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;  
 Fabre, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi,  
 A.; Galleron, N.; Ghm, S.Y.; Glaser, P.; Goffeau, A.; Goldschmidt, J.; Grand,  
 G.; Gussiepl, C.; Guy, B.J.; Hage, K.; Halech, J.; Harwood, C.R.; Hentut, A.;  
 Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.L.;  
 Kobil, B.; Karamata, D.; Kasahara, Y.; Kiebert-Blanchard, M.; Klein, C.;  
 Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.;  
 Lapidus, A.; Lardinois, S.  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda,  
 S.; Muesel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.;  
 Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega,  
 B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Portolick, S.; Prescott,  
 A.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Ray, M.; Reynolds,  
 S.; Rieger, M.; Rivolta, C.; Roeha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato,  
 T.; Scanlon, E.  
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska,  
 A.; Seror, S.J.; Serro, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.;  
 Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka,  
 T.; Terpiltra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenoel, M.;  
 Vannier, F.; Vassarotti, A.; Viari, A.; Wambuti, R.; Wedler, E.; Wedler, H.;  
 Weitenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto,  
 K.; Yata, K.; Yoshida, K.  
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus  
 subtilis.  
 A:Reference number: A69580; MUID:98044033  
 A:Accession: E70044  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-310 <KUN>  
 A:Cross-references: GB:299121; GB:AL009126; NID:g2635827; PID:e1186188;  
 PID:g2636013  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: yvob  
 C:Superfamily: hypothetical protein yvob  
 E70044 Length: 310 February 11, 2000 15:51 Type: P Check: 4650 ..

1 MAKVRKDYV EGFNLEISG EGINRPITM SDSLSPGIEI AGFTTYPRE  
 51 RVQLCKTEL SFEPQLPREE KQCRMSLCT DTPAILISR DMPIQELID  
 101 ASKNGVPLV RSPKTRLS SRLTNLESRL IAPTAIHGV LVDYGVVL  
 151 ITKSGVGRS ETALVYKRG HRLVADCV E IRQEDQTLV GNAPELIEHL  
 201 LKIRGLGIIN VMLTFGAGAV RSNKRITIVM NLELWQGRQ YDRIGLEET  
 251 MKIIDIEIKR LTIYVPRGN LAVIEVAM NFRLRMGLN ABOFTKLA  
 301 DVIDGEQEE  
 !:A SEQUENCE 1.0  
 P1:G69000 - molybdenum cofactor biosynthesis protein MoeA - Methanobacterium  
 thermoautotrophicum (strain Delta H)  
 C:Species: Methanobacterium thermoautotrophicum  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: G69000  
 R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.;  
 Aldredge, T.; Bashirzadeh, R.; Blakey, D.; Cook, R.; Gilbert, K.; Harrison,  
 D.; Hoang, L.; Keagle, P.; Lumm, W.; Potcher, B.; Qiu, D.; Spadatoro, R.;  
 Vicaltre, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso, A.;  
 Bush, D.; Sifer, H.; Patwell, D.; Prabhakar, S.; McDougall, S.; Shimer, G.;  
 Goyal, A.; Pietrovski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.;  
 Noelling, J.; Reeve, J.N.  
 J. Bacteriol. 179, 7135-7155, 1997  
 A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta  
 H: functional analysis and comparative genomics.  
 A:Reference number: A69000; MUID:98037514  
 A:Accession: G69000  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-409 <MTH>  
 A:Cross-references: GB:AE000873; GB:AE000666; NID:g2622101; PID:g2622104  
 A:Experimental source: strain Delta H  
 C:Genetics:  
 A:Gene: MTH1003  
 A:Start codon: TTG  
 C:Superfamily: molybdenum cofactor biosynthesis protein moeA-2  
 G69000 Length: 409 February 11, 2000 15:51 Type: P Check: 7972 ..

1 MNGVELEIT DLDARRITG ELFSELYSOG TESVDILDAD GRVLADIES  
 51 PVDLPPFDRA SRDGYAVRAA DFGADEBNP ATLRCIETVE AGVSPSLEVA  
 101 EGICTISTG APVEGADAV VAVEYTWEEY DRYVQRPAY PSOHIAARG  
 151 DIRAGVNLK TGYLSPDKI AALSAGITR VNVISKPSVC VISTGNELIE  
 201 PSQQLPGRKI FDSNAGLSS ALKEACSVK HGGIYRDDYG ELRDAILRGL  
 251 EKDQWITSG GTSAGAGDIL SEVLDELGRV VHGHSIKPG KPTIVGVVD  
 301 KIYVGLPGFP VSALIVPML IEPFIRKLSG RKASGHVKT LKLAERLHSS

351 RGRHYALVM VDGYPARPIT KSGGAVVALA GADGITEVPR NFEINSEEE  
401 VEVLFEENR

!!AA\_SEQUENCE 1.0  
P1:E70191 - conserved hypothetical protein BB0734 - Lyme disease spirochete  
C:Species: *Borrelia burgdorferi* (Lyme disease spirochete)  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: E70191  
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.; Dodson, R.; Hickey, E.K.; Gwin, M.; Dougherty, B.; Tomb, J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.; Kesteven, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman, J.; Uterback, T.; Wathey, L.; McDonald, L.; Artach, P.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
Nature 390, 580-586, 1997  
A:Authors: Smith, H.O.; Venter, J.C.  
A:Title: Genome sequence of a Lyme disease spirochete, *Borrelia burgdorferi*.  
A:Reference number: A70100; MUID:98065943  
A:Accession: E70191  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-337 <KLE>  
A:Cross-references: GB:AE001173; GB:AE000783; NID:92688665; PID:92688669; TIGR:BB0734  
A:Experimental source: strain B31  
C:Superfamily: *Bacillus subtilis* conserved hypothetical protein ywlc

E70191 Length: 337 February 11, 2000 15:51 Type: P Check: 8273 ..

1 MISTEISS OIOKAKLIK MGEIVFPE TVYGIGANAY NEDAVKIFL  
51 VKRRPINPL IVHVDYVKI KESEYIPKS ALMLIKRSP GPLTYVLKKS  
101 IKSRFVSGN LDTVAIRIPA NKTALSLIK SKPIVAPSA NISKRPESTN  
151 FEALKEUNG LVGIIKPEE NKDPNIGIES TVVGFDLMDN VLLIRPGAIT  
201 KMIENELQG KTVVYATK MELEKSPGN IHHKPKIPV YLRSQDNIR  
251 RYLNKOTKIL ITKATLKSYL FNFVKNKNI TVFNTLEEA QNLKELVNS  
301 ENNYKQILSE FLKDELGHS INNRKKAAS NKFINK

!!AA\_SEQUENCE 1.0  
P1:A71183 - probable UDP-glucose 4-epimerase - *Pyrococcus horikoshii*  
C:Species: *Pyrococcus horikoshii*  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: A71183  
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohnuki, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamaaki, J.; Kushiida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.  
DNA Res. 5, 55-76, 1998  
A:Title: Complete sequence and gene organization of the genome of a hyperthermophilic archaeobacterium, *Pyrococcus horikoshii* OT3.  
A:Reference number: A71000; MUID:98344137  
A:Accession: A71183  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-306 <KAW>  
A:Cross-references: GB:AP000007; NID:93236134; PID:d1031799; PID:93258173  
A:Experimental source: strain OT3  
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C:Gene: PH1742  
C:Superfamily: hypothetical protein s111213

A71183 Length: 306 February 11, 2000 15:51 Type: P Check: 8100 ..

1 MKNKLIYTG GAGFIGSHLA EALKDENDVI IIDNIXSRI ENIEPEGKFI  
51 RADYDYESI AEVISEADIV FHEAQSIVK ESEIDPTE EVANVIGINV  
101 LRAISOQDGK LIFASSAAY GEKEPLITE DTLNIPISPY GITKLAENH  
151 CRYQSILYGI PVYLIRYENV YGPRQSSAVA GVISIFLERA IKGEPLIIFG  
201 DKGOTRPFIV YKDVEANIL VAKKRSANGR IENVATGKET TIELANKII  
251 DMTSSSSSIL FYPFPGDIR RSVAKIERIK KLGFRPRSL EELKETEFKW  
301 FTSRTQ

!!AA\_SEQUENCE 1.0  
P1:S70955 - ctnf protein - *Vibrio cholerae*  
C:Species: *Vibrio cholerae*  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: S70955  
R:Balk, E.M.; Bunschoten, A.E.; Willems, R.J.L.; Chang, A.C.Y.; Mool, F.R.  
Mol Microbiol 20, 799-811, 1996  
A:Title: Genetic organization and functional analysis of the ctn DNA essential for cell wall polysaccharide synthesis in *Vibrio cholerae* O139.  
A:Reference number: S70952; MUID:96586047  
A:Accession: S70955  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-288 <BKR>  
A:Cross-references: EMBL:X90547; NID:91462276; PID:e194952; PID:g1107920  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995  
C:Gene: ctnf  
C:Superfamily: hypothetical protein b0985

S70955 Length: 288 February 11, 2000 15:51 Type: P Check: 8920 ..

1 MCGTDLFRL RSPFKICRSP IKKRAWSFIT SSGWITLAAA ISGSLCVA  
51 PLATSEAVR ANNOAIST RLSPFEPNRL DSAVQTLQ AGLTQOVEM  
101 PSAGLFLSH AFLFKRDVLL KLAQOQSSAP PQOALMAST IAQROAEFA  
151 KRLFISVDP WRIAPQHP RLNSWMLTL NKSQTQVSIV GAVNQPDVI  
201 WHNLSAKDY ARAAGLIDQ ISEIVIQPD GIAOKHAYV WNOFNEVAP  
251 GAIYVPLPL KRAFPDPTV DADLNQVIE LERNRLPL

!!AA\_SEQUENCE 1.0  
P1:A70118 - glycine betaine, L-proline ABC transporter, permease protein (prow) homolog - Lyme disease spirochete  
C:Species: *Borrelia burgdorferi* (Lyme disease spirochete)  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: A70118  
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.; Dodson, R.; Hickey, E.K.; Gwin, M.; Dougherty, B.; Tomb, J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.; Kesteven, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman, J.; Uterback, T.; Wathey, L.; McDonald, L.; Artach, P.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
Nature 390, 580-586, 1997  
A:Authors: Smith, H.O.; Venter, J.C.  
A:Title: Genome sequence of a Lyme disease spirochete, *Borrelia burgdorferi*.  
A:Reference number: A70100; MUID:98065943  
A:Accession: A70118  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-299 <RLE>  
A:Cross-references: GB:AE001125; GB:AE000783; NID:926886021; PID:926886023;

TIGR:BB0145  
 A:Experimental source: strain B31  
 C:Superfamily: glycine betaine/carnitine/choline ABC transporter  
 A70118 Length: 299 February 11, 2000 15:51 Type: P Check: 6242 ..

1 MTKDFEIKTI DNEFDELVDN FSISDVGFS KSIIFFLENT KMLFFVNF  
 51 LEFIVGLLS EYFLKKRLIE ILPGFEFLL YFNLMESMD TIAIFVSVL  
 101 VSVILGIPIG ILGCEPPEY VFLKPIIDLM QAMPPEYILL PAIPFEGMGT  
 151 ASAIFATIVF AMPPIRYTR LGIVQVSDV IEAKSPGSS NRIIFQVOL  
 201 PLISQSIIEG INOSIMMAIS MIVIAWVS SGLGRVIYS IERLNGEGL  
 251 ISGLAVYIA IILDRIMQSI FIKFSYLTND HYGKKEKNEF KRFLLEYNK

!!AA\_SEQUENCE 1.0  
 P1:S76375 - plastocyanin precursor - Synechocystis sp. (strain PCC 6803)  
 N:Alternate names: protein sl10139  
 C:Species: Synechocystis sp.  
 A:Variety: PCC 6803  
 C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 11-Jun-1999  
 C:Accession: S76375; S13733  
 R:Kaneh, T.; Sato, S.; Kocani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi, I.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naito, K.; Okumura, S.; Shimpo, S.; Tanouchi, C.; Wada, I.; Metanabe, A.; Yamada, M.; Tasuda, M.; Tabata, S.  
 DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.  
 A:Reference number: S74322; MUID:97061201  
 A:Accession: S76375  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-126 <EAN>  
 A:Cross-references: EMBL:D64000; GB:AB001339; NID:g1001484; PIDN:BA010227.1; PID:d1010878; PID:g1001599  
 A:Experimental source: PCC 6803  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

R:Briggs, L.M.; Pecoraro, V.L.; McIntosh, L.  
 Plant Mol. Biol. 15, 633-642, 1990  
 A:Title: Copper-induced expression, cloning, and regulatory studies of the plastocyanin gene from the cyanobacterium Synechocystis sp. PCC 6803.  
 A:Reference number: S13733; MUID:91338701  
 A:Accession: S13733  
 A:Molecule type: DNA  
 A:Residues: 1-126 <BR1>  
 A:Cross-references: EMBL:X54105; NID:g47401; PIDN:CAA38038.1; PID:g47402  
 A:Experimental source: PCC 6803  
 R:Romero, A.; De La Cerdá, B.; Varela, P.F.; Navarro, J.A.; Hervás, M.; De La Rosa, M.A.  
 Submitted to the Brookhaven Protein Data Bank, June 1997  
 A:Reference number: A68539; PDB:1PCS  
 A:Contents: annotation; X-ray crystallography, 2.15 angstroms, 29-71, 'D', 73-76, 'P', 78-89, 'L', 91-126  
 A:Note: engineered sequence expressed in Escherichia coli  
 C:Genetics:  
 A:Gene: petE  
 C:Superfamily: plastocyanin  
 C:Keywords: copper; electron transfer; metalloprotein  
 F:1-18/Domain: signal sequence #status predicted <SID>  
 F:19-28/Domain: propeptide #status predicted <PRO>  
 F:29-126/Product: plastocyanin #status predicted <MAT>  
 F:67,111,114,119/Binding site: copper (His, Cys, His, Met) (type 1) #status experimental

S76375 Length: 126 February 11, 2000 15:51 Type: P Check: 6628 ..

1 MSKFLITILA GLLVVSFF LSVSPAAN ATYKMSDSC ALVEPSTVT

51 IKAGEYKRV NKLSPHNIV FAADGVADT AAKISHKGLA FAAGESFTST  
 101 FTEPETYTYT CEPHRCAGMY GKAYVE

!!AA\_SEQUENCE 1.0  
 P1:F70646 - probable quinine oxidoreductase showing strong - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 26-Aug-1999  
 C:Accession: F70646  
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churche, C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: F70646  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-323 <COL>  
 A:Cross-references: GB:283867; GB:AL123456; NID:g3261695; PID:e291013; PID:g18126  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: fadB4  
 C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology  
 F:23-310/Domain: long-chain alcohol dehydrogenase homology <LAOH>

F70646 Length: 323 February 11, 2000 15:51 Type: P Check: 8411 ..

1 MRAVVRTRLE GPDAVEVAEY EEPISAGYVI EVNAGVAFP DALLTRGRYO  
 51 YRPEPFVGL AEIAGVYRSA PONSQVRSQD RVVGLTALTG GMAEVAVLSP  
 101 ERFVFLPDNM TFEAGAGVLF NDLVYFALA VRGRLQGET VLVHGAAGSI  
 151 GSTILRLAPA LGASRTVAVY STQEKRELAT VAGATDVLA EEFKDAVOEL  
 201 TNGSGVDIVY DVGSGRFTD SLRSLAAGR LTVIGFTGGE IPTVKVNRLL  
 251 LNNIDVYGVG WGAWSLTHPD ALAQWQSOLE RLRSCKLPP PEPVYPIPDQ

301 AAAAIAISEN RTAKKRVLR VRD

!!AA\_SEQUENCE 1.0  
 P1:S01901 alcohol dehydrogenase (EC 1.1.1.1) 2 - fruit fly (Drosophila melanogaster)  
 C:Species: Drosophila melanogaster  
 C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 11-Jun-1999  
 C:Accession: S01901  
 R:Atkinson, P.W.; Mills, L.E.; Starmer, W.T.; Sullivan, D.T.  
 Genetics 120, 713-723, 1988  
 A:Title: Structure and evolution of the Adh genes of Drosophila melanogaster.  
 A:Reference number: S01901; MUID:89137903  
 A:Accession: S01901  
 A:Molecule type: DNA  
 A:Residues: 1-254 <ATK>  
 A:Cross-references: EMBL:X12536; NID:g7553; PIDN:CAA31054.1; PID:g7554  
 C:Genetics:  
 A:Gene: Adh-2  
 A:Cross-references: FlyBase:FBgn0012567  
 A:Introns: 31/3; 166/3  
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology  
 C:Keywords: alcohol metabolism; NAD; oxidoreductase  
 F:6-182/Domain: short-chain alcohol dehydrogenase homology <SADH>

S01901 Length: 254 February 11, 2000 15:51 Type: P Check: 8887 ..

1 MAIANKNIF VAGLGIGFD TSREIVKSGP KNVLIDRIE NPAAIELEKA

51 LNFRTVTFE PYDVTVSVAE TTKLKTIIF KIKTVLDLIN GTGILDDHOI

101 ERTIAVETG TLNTTAIMS FMDKRGCGP GTIANICSVT GFNALLPVV

151 YSASKAAALS FTNSLARIAP ITGVAYSIN PGITTTILVH KFNGLDVEP

201 RVALLLEHP TOTLQCAON FVAKIQANON GAIWMLDGT LEALEWTKKH

251 DSHI

!!AA\_SEQUENCE 1.0

F1:S65732 - carbonyl reductase (NADPH) (EC 1.1.1.184) - chicken (fragments)

C:Species: Gallus gallus (chicken)

C>Date: 06-Dec-1996 #sequence\_revision 13-Mar-1997 #text\_change 07-Aug-1998

C:Accession: S65732

R:Tsal, S.P.; Wang, L.Y.; Yeh, H.I.; Tam, M.F.

Biochim. Biophys. Acta 1292, 288-292, 1996

A:Title: The 30 kDa protein co-purified with chick liver glutathione S-transferases is a carbonyl reductase.

A:Reference number: S65732; MUID:96176857

A:Accession: S65732

A:Status: Preliminary

A:Molecule type: protein

A:Residues: 1-13;14-42;43-58;59-105;106-127;128-132 <TSA>

C:Superfamily: Ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

C:Keywords: oxidoreductase

F:4-35/Domain: short-chain alcohol dehydrogenase homology (fragments) <SADH>

S65732 Length: 132 February 11, 2000 15:51 Type: P Check: 4453 ..

1 SNVPAVVTG TNKPXGQEA VAKLOEGLH PLFHOLDID LQVSDRTPFA

51 VOAEVTLKGV HEKEGMPNSA YGVSKIGTV LSRIOARMAN ESKRGHILL

101 NAXXPAPKSP EGAETPVYL ALPPDAIVR TW

!!AA\_SEQUENCE 1.0

P1:JC5284 - carbonyl reductase (NADPH) (EC 1.1.1.184), Inducible - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 16-Apr-1997 #sequence\_revision 18-Jul-1997 #text\_change 11-Jun-1999

C:Accession: S68983; PC2334; JC5284; PC2159; S52349

R:Wernuth, B.; Maeder-Heinmann, G.; Ernst, E.

Eur. J. Biochem. 228, 473-479, 1995

A:Title: Cloning and expression of carbonyl reductase from rat testis.

A:Reference number: S68983; MUID:95220378

A:Accession: S68983

A:Status: Preliminary

A:Molecule type: mRNA

A:Residues: 1-277 <WER>

A:Cross-References: EMBL:X84349; NID:9666086; PIDN:CAA59088.1; PID:9666087

R:Rolt, E.; Soederstrom, M.; Ahlberg, M.B.; DePierre, J.W.

Biochem. Biophys. Res. Commun. 201, 149-154, 1994

A:Title: A novel 34kDa glutathione-binding protein in mature rat ovary.

A:Reference number: PC2159; MUID:94256971

A:Accession: PC2234

A:Molecule type: protein

A:Residues: 104-121,'X',123-134,'D',136-137,'H' <TOF>

A:Experimental source: ovary

R:Aoki, H.; Okada, T.; Mizutani, T.; Numata, Y.; Mitsuhashi, T.; Miyamoto, K.

Biochem. Biophys. Res. Commun. 230, 518-523, 1997

A:Title: Identification of two closely related genes, inducible and noninducible carbonyl reductases in the rat ovary.

A:Reference number: JC5284; MUID:97167735

A:Contents: ovary

A:Accession: JC5284

A:Molecule type: mRNA

A:Residues: 1-140,'GW',143,'R',145-235,'T',237-238,'E',240-277 <AK>

A:Cross-References: DDBJ:DB9069; NID:9106811; PIDN:BAI19007.1; PID:91019750; PID:91906812

C:Comment: This enzyme is a cytosolic, monomeric oxidoreductase that catalyzes the NADPH-dependent reduction of variety of endogenous carbonyl compounds.

C:Genetics:

A:Gene: 1CR

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

C:Keywords: cytosol; monomer; NADP; oxidoreductase

F:6-185/Domain: short-chain alcohol dehydrogenase homology <SADH>

JC5284 Length: 277 February 11, 2000 15:51 Type: P Check: 7185 ..

1 MSSDRPALV TGANKSIGFA IVRDLCKREL GDVVLARDE SRGHEAVROL

51 QTEGLSPRFH QLDIDNPQI RALRDFLQE YGGLNLYVN AGIAFKVVD

101 TPFHQAEVT MKNTEFGTD VCKELLPILK POGRYVNVSS SVSLRLKSC

151 SPELOKERS ETITEELVG LMKFIEDPAK KGVHAKGWP NSAYGVTKIG

201 VTLRSRIYAR KLINEEREDK ILLNACPGW VRTDMAGPKA TKSPEGAE

251 PVYIALIPG AEGPHQGFVQ DKVEPW

!!AA\_SEQUENCE 1.0

P1:S54815 - undamycin polyketide ketoreductase (EC 1.1.1.-) - Streptomyces fradiae

C:Species: Streptomyces fradiae

C>Date: 08-Jul-1995 #sequence\_revision 21-Jul-1995 #text\_change 11-Jun-1999

C:Accession: S54815

R:Decker, H.; Haag, S.

submitted to the EMBL Data Library, May 1995

A:Description: Cloning and characterization of a polyketide synthase from Streptomyces fradiae 9ue217 encoding the genes for biosynthesis of the undamycin antibiotic undamycin A and a gene probably involved in its oxygenation

A:Reference number: S54810

A:Accession: S54815

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-261 <DEC>

A:Cross-References: EMBL:X87093; NID:9809102; PIDN:CAA60572.1; PID:9809108

C:Superfamily: Ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

C:Keywords: NAD; oxidoreductase

F:7-188/Domain: short-chain alcohol dehydrogenase homology <SADH>

S54815 Length: 261 February 11, 2000 15:51 Type: P Check: 5190 ..

1 MTOONPKVAL VTGATSGIGL AVTRLLGRQG HRFVLCARTE DSVSTVKOL

51 LDEGLEVDGA PCDVRSADVD ERFVORAVHR FETIDVLVNN AGREGGVTA

101 DIADELMHVDY IDTNLSVFR MTRVINTGS MRHKDRGRII NINASTACKOG

151 VVLGAPYSAS KHGVVGFITKA LGNELAPITGI TYNAVCPQIV ETPMAQNVRA

201 GYVAAADTSE DALLEKPAK IPIGRSTPE EVAGLVGVLA SDTAASTTSO

251 ALNVCGGLGN F

!!AA\_SEQUENCE 1.0

P1:D5587 - grieseus polyketide ketoreductase (EC 1.1.1.-) - Streptomyces grieseus

C:Species: Streptomyces grieseus

C>Date: 25-Aug-1995 #sequence\_revision 25-Aug-1995 #text\_change 11-Jun-1999

C:Accession: D5587

R:Yu, T.W.; Bibb, M.J.; Revill, W.P.; Hopwood, D.A.

J. Bacteriol. 176, 2627-2634, 1994

A:Title: Cloning, sequencing, and analysis of the grieseus polyketide synthase gene cluster from Streptomyces grieseus.

A:Reference number: A55587; MUID:94222841

A:Accession: D5587

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-261 <YUA>

A:Cross-references: GB:X77865; NID:G488770; PIDN:CAA54861.1; PID:G488774  
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology  
C:Keywords: NAD; oxidoreductase  
F:7-188/Domain: short-chain alcohol dehydrogenase homology <SADH>  
D5587 Length: 261 February 11, 2000 15:51 Type: P Check: 3920 ..

1 MSDODRVAV VTGATSGIGL SVARLLASOG HRVFIGARSA ENVAATKEL  
51 VAEGLDVDA VDVRSADV EAWIOAAVDR FGTVDVYVNN AGRSSGGVTA  
101 DIDPELMDV IDTNLSVFR VTRSAALNTGG LRSKSRGRII NIASTGKOG  
151 VVLGAPYSAS KHGVGFTKA LGNELAPTGI TVNAVCPGY ETPMAQRVQ  
201 GYAAAYDATE DAILEKFOAK IPLGRYSTPD EVAGLVGYLA SDTAASITSQ  
251 ALNVCGLGN F

!!AA\_SEQUENCE 1.0  
P1:S69225 - nogalamycin polyketide ketoreductase (EC 1.1.1.-) snod -  
Streptomyces nogalater  
C:Species: Streptomyces nogalater  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 17-Mar-1999  
C:Accession: S69225  
R:Ylthorke, K.; Tuikka, J.; Jussila, S.; Cong, L.; Meentsaelae, P.  
Mol. Gen. Genet. 251, 113-120, 1996  
A:Title: A gene cluster involved in nogalamycin biosynthesis from Streptomyces  
nogalater: sequence analysis and complementation of early-block mutations in  
the anthracycline pathway.  
A:Reference number: S69223; MUID:96242142  
A:Accession: S69225  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-262 <YLI>  
A:Cross-references: EMBL:Z48262; NID:G1181156; PID:G221092; PID:G1181158  
A:Experimental source: ATCC 27451  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library,  
February 1995  
C:Genetics:  
A:Gene: snod  
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology  
C:Keywords: NAD; oxidoreductase  
F:7-189/Domain: short-chain alcohol dehydrogenase homology <SADH>  
S69225 Length: 262 February 11, 2000 15:51 Type: P Check: 7076 ..

1 MPTDGRVAV VTGATSGIGL AVYKALAATG YACRSILCAR SDAVDSTVEE  
51 LRDEKLEVDG RSCDVSTDD VAEFTGFATR RFGGVYLVN NGRNGGGPT  
101 ARITDELMD VIEITNLTSVF RVTREVLVTG GMLAAGRRI INIATSGKQ  
151 GVVLGAPYSA SKHGVGFTK ALGLELAKSG ITNAVCPGY VETPAGERV  
201 RGYAEVMTS EEAVERFOA KIPLEGVAPQ EYVAGLVTYL TPTADISITA  
251 QALNVCGLGN Y

!!AA\_SEQUENCE 1.0  
P1:S35196 - hypothetical protein 3 - Saccharopolyspora hirsuta  
C:Species: Saccharopolyspora hirsuta  
C:Date: 10-Dec-1993 #sequence\_revision 21-Jul-1995 #text\_change 28-May-1999  
C:Accession: S35196  
R:Le Gouill, C.; Desmarais, D.; Dery, C.V.  
Mol. Gen. Genet. 240, 146-150, 1993  
A:Title: Saccharopolyspora hirsuta 367 encodes clustered genes similar to  
ketoadl synthase, ketoadl reductase, acyl carrier protein, and biotin  
carboxyl carrier protein  
A:Reference number: S35194; MUID:93341453  
A:Accession: S35196  
A:Status: Preliminary  
A:Molecule type: DNA

A:Residues: 1-261 <GON>  
A:Cross-references: GB:M88258; NID:G347176; PIDN:AAA26487.1; PID:G347179  
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology  
C:Keywords: NAD; oxidoreductase  
F:7-188/Domain: short-chain alcohol dehydrogenase homology <SADH>  
S35196 Length: 261 February 11, 2000 15:51 Type: P Check: 9329 ..

1 MAANTPVAV VTGATSGIGL ATRRLGQUG ARVTCARRA EELATITKIL  
51 QAEGLVDGS TCDVSVPGQI PELVAAVDR FEPIDVLVN AGRSSGGVTA  
101 EYDETFEDV INTNLNSVFL MTKOVLTKG MLEDERGRII NIATSGKOG  
151 VVNGSPYSAS KHGVGFTKS LGLELAKTGI TVNAVCPGY ETPMAQRVQ  
201 HYSKIVSSE AEADRVTR VPLGRVPEE EVAAVEYLV SDGAAAVTAQ  
251 ALNVCGLGN Y

!!AA\_SEQUENCE 1.0  
P1:D64051 - 3-oxoacyl-[acyl-carrier-protein] reductase (EC 1.1.1.100) -  
Haemophilus influenzae (strain Rd KW20)  
C:Species: Haemophilus influenzae  
C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 11-Jun-1999  
C:Accession: D64051  
R:Kielbaso, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.;  
Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.; Dougherty, B.A.; Merrick, J.M.;  
McKenney, K.; Sutton, G.; Fitzhugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;  
Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton, M.D.; Utterback, T.R.; Hanna, M.C.;  
Nguyen, D.T.; Saudek, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.;  
Fuhman, J.L.; Geophagen, N.S.M.  
Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.;  
Venter, J.C.  
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae  
Rd.  
A:Reference number: A64000; MUID:95350630  
A:Accession: D64051  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-242 <TIGR>  
A:Cross-references: GB:U22701; GB:L42023; NID:G1573105; PIDN:AAC21824.1;  
PID:G157312; TIGR:H10135  
C:Genetics:  
A:Gene: tabg  
C:Function:  
A:Pathway: fatty acid biosynthesis  
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology  
C:Keywords: fatty acid biosynthesis; NADP; oxidoreductase  
F:4-180/Domain: short-chain alcohol dehydrogenase homology <SADH>  
F:4-34/Region: beta-alpha-beta NADP nucleotide-binding fold  
F:149/Active site: Tyr #status predicted  
D64051 Length: 242 February 11, 2000 15:51 Type: P Check: 4159 ..

1 MOKIALVTG STRIGRAIA EELSKGAFV IGTATSEKA EATSAVLGDK  
51 GKGLVAVTD KESITLLEQ IKNDFDIDI LVNNAGITRD NLIMRKKDEE  
101 WEDIMQNTT SVYHLSKRL RSMKKRFGR IINIGSVGS TGNPGQTNVC  
151 AAKAGVGS KSLAKEVAR GITVNVAVG FIANDMEVL TDQRKGILS  
201 NVPAGRLGEA KDIKAVAFV ASDDAVYIIG TLIHNGSLY LS

!!AA\_SEQUENCE 1.0  
P1:D56648 - 2-keto-3-deoxygluconate oxidoreductase kdud - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 11-Jun-1999  
C:Accession: D56648  
R:Kunst, F.; Ogatawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo,  
V.; Bertero, M.G.; Besteres, F.; Boiotin, A.; Borchert, S.; Boriss, R.;



Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denzot, F.; Devine, R.M.; Dietzenhoef, A.; Ehrlich, S.D.; Emerson, P.T.; Entlan, K.D.; Errington, J.; Fabbet, C.; Ferrari, E.  
 Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Galleon, N.; Gilm, S.Y.; Glaser, P.; Goffeau, A.; Goldthly, E.; Grandi, G.; Guisepi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Klein, C.; Joris, B.; Karamata, D.; Kashara, Y.; Kieffer-Bianhard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, A.; Authors: Lauber, U.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Potwolk, S.; Prescott, A.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Ray, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.  
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Seror, P.; Shin, B.S.; Soldo, B.; Sotokhin, A.; Taccini, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tanaka, A.; Tanaka, T.; Terstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandebol, M.; Vannier, F.; Vassarotti, A.; Varti, A.; Wandt, R.; Wedler, E.; Wedler, H.; Weitenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
 A:Reference number: A69580; MUID:98044033  
 A:Accession: D69648  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1254 <UN>  
 A:Cross-references: GB:Z99115; GB:AL009126; NID:92634478; PID:CB1431.1; FID:118360; FID:82634633  
 A:Experimental source: strain 168  
 A:Genetics:  
 A:Gene: Kudu  
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology F:13-190/Domain: short-chain alcohol dehydrogenase homology <SADH>

D69648 Length: 254 February 11, 2000 15:51 Type: P Check: 3484 ..

1 MGYLHDAFSL KGTALVTGP GTGIGGIAK ALAGAGADI GTSHTSLSE  
 51 TQQLVEDEGR IFTSTLDS KPAIKDSAA ELFENQIDI LYNNAGIHR  
 101 EKADPEEN WOHVNVNIN SLFITOLAG RMLKRGHR IINIASLSF  
 151 OGGLVPAVT ASKRAVAGIT KSFANEMAS GIOVNAIDAG YISTANKPI  
 201 RDEKRNEDI LKRIPAGRW QADDIGTAV FLASRASDVY NGHILAVDG  
 251 WLSR

!!AA\_SEQUENCE 1.0  
 C:Species: Datura stramonium (jimsonweed, common thornapple)  
 C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 11-Jun-1999  
 C:Accession: C48674  
 R:Nakajima, K.; Hashimoto, T.; Yamada, Y.  
 Proc. Natl. Acad. Sci. U.S.A. 90, 9591-9595, 1993  
 A:Title: Two tropione reductases with different stereospecificities are short-chain dehydrogenases evolved from a common ancestor.  
 A:Reference number: A48674; MUID:94022421  
 A:Accession: C48674  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-268 <NAK>  
 A:Cross-references: GB:U20475; NID:9424157; PID:AAA33280.1; PID:9424158  
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

F:18-198/Domain: short-chain alcohol dehydrogenase homology <SADH>  
 C48674 Length: 268 February 11, 2000 15:51 Type: P Check: 5293 ..

1 MAGREIIGGD RWSLRGNTA LVYGTGRGIG YAIVEELANF GAEVITCSRS  
 51 QNDDECLEK WRKGFVSG PCDVSSISO ROTIMESVTS SENGKINILI  
 101 NNATTIPKE ATNFTAEYS IIMGTFEAS YNLCLAHPL LKASGNASIV  
 151 FNSAAAVIA VLSSIVAS KAINQVTKS LACEWAKDSI RVNAVAPWII  
 201 NPIPEACQ VPSQKNIES LIGRAPMKRA GPSEVSSLV TYICLPTASY  
 251 ITGOLICVDG GTYVNGFI

!!AA\_SEQUENCE 1.0  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 20-Sep-1999  
 C:Accession: T12020  
 R:Kutchna, A.J.; Hoang, T.T.; Schweizer, H.P.  
 submitted to the EMBL Data Library, February 1997  
 A:Description: Molecular characterization of the acpP-containing region of the *Pseudomonas aeruginosa* chromosome.  
 A:Reference number: 217380  
 A:Accession: T12020  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-247 <KUT>  
 A:Cross-references: EMBL:U91631; NID:92738151; PID:92738155  
 C:Genetics:  
 A:Gene: fabG  
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology T12020 Length: 247 February 11, 2000 15:51 Type: P Check: 9922 ..

1 MSIGKVALY TGASRGIGQ IALELGRUGA VVIGTATSAS GAETIAETLK  
 51 ANGEVAGLV LDVSSDESA ATLEHQHQL GQPLIVVNA GITRDNLLVR  
 101 MKDDEFDVY NTNINSLXRL SKAVLRGMR ARWGRINIG SYVAGMGNG  
 151 QNVAANKAG LEGTFLAR EVGSRATVN AVAGFIDTD MRELEPEAOR  
 201 EALIGQIPUG RLGQAEIYK VVGFLASDGA AVVGTATVP NGMWYMS

!!AA\_SEQUENCE 1.0  
 C:Species: *Vibrio harveyi*  
 C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 20-Sep-1999  
 C:Accession: T12051  
 R:Shen, Z.; Byers, D.M.  
 J. Bacteriol. 178, 571-573, 1996  
 A:Title: Isolation of *Vibrio harveyi* acyl carrier protein and the *fabG*, *acpP* and *fabF* genes involved in fatty acid biosynthesis.  
 A:Reference number: 217396; MUID:96134997  
 A:Accession: T12051  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-244 <SHD>  
 A:Cross-references: EMBL:U39441; NID:91173839; PID:91173841  
 A:Experimental source: strain B392  
 A:Genetics:  
 A:Gene: fabG  
 C:Function: fatty acid biosynthesis  
 A:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology  
 C:Keywords: fatty acid biosynthesis; MDP; oxidoreductase  
 T12051 Length: 244 February 11, 2000 15:51 Type: P Check: 5984 ..

1 MNIEGKIALV TGASRGICRA IAEVLVERGA TVIGRTASEG GAANAISHTLG  
 51 ENGKGLALNV TDVESIEATL KTIINDECGAI DILVNNACIT RDLNLMKMD  
 101 DEMNDIINTN LTPYRMSKA VLKGMKKRA GRINWGSV GTMGNAQOIN  
 151 YAAKAGVIG FTKSMAREVA SRGVYNTVA POFIETDMTK ALNDQRAAT  
 201 LSNVPAERLG DPREIASAVV FLASPEAAV TGETLHVNG MYWV

!!AA.SEEQUENCE 1.0  
 P1:S39654 - hypothetical protein X - Pseudomonas aeruginosa  
 C:Species: Pseudomonas aeruginosa  
 C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 26-Aug-1999  
 C:Accession: S39654  
 R:Krim, M.; Bally, M.; Ball, G.; Tommassen, J.; Teerink, H.; Filloux, A.; Lazdunski, A.  
 MO: Microbiol. 10, 431-443, 1993  
 A:Title: Acp-mediated protein secretion in Pseudomonas aeruginosa: identification of two additional genes and evidence for regulation of xcp gene expression.  
 A:Reference number: S39652; MUID:95020542  
 A:Accession: S39654  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-255 <AKR>  
 A:Cross-references: EMBL:X68594; NID:9431183; PID:9581439  
 C:Genetics:  
 A:Start codon: GTG  
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology F:11-187/Domain: short-chain alcohol dehydrogenase homology <SADH>  
 S39654 Length: 255 February 11, 2000 15:51 Type: P Check: 5853 ..

1 MIEPPIRNGG KVALYTGAR GIGLGISANL IAEQGVVLA DNDREKARY  
 51 AEALEGHANF VAMDYAQEQ VANSVAEVLG QFGRIDGLVC MAIAPRNT  
 101 PLEALSIGEM TRTIAVNTLG PMLAKYCTP YIRAHNGAIV NIASTRHQ5  
 151 EPDSKAYNAAS KGLIALTHA LAASLGPDIR VNALSPGMD TREAREEA  
 201 PLEIHDHDOH LVGRVYVED VASLVAMLS EDAGFVTQGE FLYDGMTRK  
 251 MYID

!!AA.SEEQUENCE 1.0  
 P1:H69885 - 3-oxoacyl- acyl-carrier protein reductase homolog ymfI - Bacillus subtilis  
 C:Species: Bacillus subtilis  
 C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 26-Aug-1999  
 C:Accession: H69885  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Bertero, M.G.; Bessières, P.; Bolochin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignelli, S.C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denzot, F.; Devine, K.M.; Duesterhoft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleon, N.; Ghm, S.Y.; Glaeser, P.; Goffeau, A.; Goldlighty, E.J.; Grandt, G.; Guiser, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetler, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Priesen, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.  
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Solio, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takekura, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandemol, M.; Vanlier, F.; Vassartelli, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Welteregger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
 A:Reference number: A69580; MUID:98044033  
 A:Accession: H69885  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-242 <KUN>  
 A:Cross-references: GB:299112; GB:AL009126; NID:92633902; PID:e1185278; PID:92634059  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: ymfI  
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology F:3-181/Domain: short-chain alcohol dehydrogenase homology <SADH>  
 H69885 Length: 242 February 11, 2000 15:51 Type: P Check: 6790 ..

1 MNKTLITGA SCGIGKISE TLAARGVNL LHTNTQNNAA AELAEKLSQM  
 51 FGVAKEIQA DLSAQDADR LTSSIVCPID AIVLNGRSH EGLIDVDNA  
 101 TVQEWQOHV ASPYMLTRNL LPMIRNKG AIYAWSING EHGACEVLY  
 151 SMAGQACISF VKLAKELAP SGIRNAYAP GAVDTMMNG FTAPAKKEIA  
 201 DEIPRGRGR TORIADATAP LISEKASYIT GOLLSNGMH HC

!!AA.SEEQUENCE 1.0  
 P1:H69868 - glucose 1-dehydrogenase homolog ykwo - Bacillus subtilis  
 C:Species: Bacillus subtilis  
 C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 26-Aug-1999  
 C:Accession: F69868  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Bertero, M.G.; Bessières, P.; Bolochin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignelli, S.C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denzot, F.; Devine, K.M.; Duesterhoft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleon, N.; Ghm, S.Y.; Glaeser, P.; Goffeau, A.; Goldlighty, E.J.; Grandt, G.; Guiser, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetler, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Priesen, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.  
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Solio, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takekura, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandemol, M.; Vanlier, F.; Vassartelli, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Welteregger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

Mon Feb 14 08:07:32 2000

plr.cat

Page 71

A:Reference number: A69580; MUID:98044033  
A:Accession: F69868  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-248 <RDN>  
A:Cross-references: GB:Z99111; GB:AL009126; NID:g2633699; PID:e1184967;  
PID:g2633748  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: ykvo  
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology  
F:7-182/Domain: short-chain alcohol dehydrogenase homology <SADH>  
F69868 Length: 248 February 11, 2000 15:51 Type: P Check: 7200 ..

1 MGKEGKIAL VTGTSIGL ATAQKRVNG AYYITGRQ NELDRAVQI  
51 GKNTGYOGD ISKLELDKL YDIKQKQK LDILFANAGI GNFPLGEIT  
101 EEOVDREFDI NVKGIETVQ KALSLFPDKV GSIIITGSTA GSIGNPASFV  
151 YGASKAALRA LVKNWILDK GTEIRNVVS PGGILTPAYD EFGDALSBEV  
201 LENSRTVPA GKVTPEEVA NAVSFLASDE SSYLTVLEL VDGILAQV

!!AA\_SEQUENCE 1.0  
P1:B55850 - hypothetical protein (hsch 5' region) - Clostridium sordellii  
C:Species: Clostridium sordellii  
C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 26-Aug-1999  
C:Accession: B55850  
R:Colman, J.P.; Hudson, L.L.; Adams, M.J.  
J. Bacteriol. 176, 4865-4874, 1994  
A:Title: Characterization and regulation of the NADP-linked  
gamma-hydroxyesteroid dehydrogenase gene from Clostridium sordellii.  
A:Reference number: A55850; MUID:9437451  
A:Accession: B55850  
A:Molecule type: DNA  
A:Status: preliminary  
A:Residues: 1-267 <COL>  
A:Cross-references: GB:112058; NID:g431280; PID:g431281  
A:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology  
F:7-189/Domain: short-chain alcohol dehydrogenase homology <SADH>  
B55850 Length: 267 February 11, 2000 15:51 Type: P Check: 1047 ..

1 MNKLEKVAL VTSAPRGICL ASAIRKLAONG AIYMGVRL EATQETCDKY  
51 KEELILKPV FPDVNIIDIT KEMIDTIKN EKKIDILVNN PETGPEKDL  
101 DLVNGEDDTF PELFNVNGS VYRLSKLIIP HMIENKGGI VNISVGGSI  
151 PDIRIRIGV SKSGVNITK QIAIOYAKG IRCNAVLPGL IATDAANSM  
201 PDERRKFLS HVLPLRIGNP EDIANSVLEF VPESDSYIT GSILEVSGCY  
251 NLGTPQIAEF VGSKYVE

!!AA\_SEQUENCE 1.0  
P1:A56275 - 1,3-propanediol dehydrogenase (EC 1.1.1.202) - Citrobacter freundii  
C:Species: Citrobacter freundii  
C:Date: 03-Oct-1995 #sequence\_revision 03-Oct-1995 #text\_change 11-Jun-1999  
R:Daniel, R.; Bosenick, R.; Gottschalk, G.  
J. Bacteriol. 177, 2151-2156, 1995  
A:Title: Purification of 1,3-propanediol dehydrogenase from Citrobacter  
freundii and cloning, sequencing, and overexpression of the corresponding gene  
in Escherichia coli.  
A:Reference number: A56275; MUID:95238288  
A:Accession: A56275  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-387 <DAN>  
A:Cross-references: GB:U09771; NID:g1229153; PIDN:AB4848.1; PID:g493086

C:Genetics:  
A:Gene: dhdt  
C:Superfamily: lactaldehyde reductase; lactaldehyde reductase homology  
C:Keywords: homocitramine; oxidoreductase  
F:8-383/Domain: lactaldehyde reductase homology <LAR>  
A56275 Length: 387 February 11, 2000 15:51 Type: P Check: 101 ..

1 MSYRMDIYLV PRVNFEPNA ISVGERCKL LGGRKALLVT DKGLRAIKDS  
51 AVDKITLHLR EAGIDVYVFD GVEPNKRDN VADGLEVRK EHCIDIYTVG  
101 GGSPPDCGKG IGIATHEGD LYSYAGIELT TNPUPPIYAV NTAGTASEV  
151 TRHCVLNTR TKRVFIYSW RNLPSVIND PLMLGKRAP LTAATGMDAL  
201 THAVEYISK DANPTDAAA IQAIRLIARN LRQAVALGSN LKARENNAVY  
251 SLAGMAFNN ANLGVHAMA HÖLGGLYDMP HGVANAVLLP HVARNYNIAN  
301 PEKADIAEF MGENTDGLST MDAEELAIHA IARLSADIGI PÖHLRDLGVK  
351 EADFPYMAEM ALKQDNAFSN PRKGNREKTA EIFRQAF

!!AA\_SEQUENCE 1.0  
P1:S47829 - glycerol-3-phosphate dehydrogenase (NAD+) (EC 1.1.1.8) -  
Escherichia coli  
C:Species: Escherichia coli  
C:Date: 27-Jan-1995 #sequence\_revision 27-Jan-1995 #text\_change 11-Jun-1999  
C:Accession: S47829; B65161  
R:Plunkett, G.  
Submitted to the EMBL Data Library, March 1994  
A:Reference number: S47829  
A:Accession: S47829  
A:Molecule type: DNA  
A:Status: preliminary  
A:Residues: 1-339 <PLU>  
A:Cross-references: EMBL:U00039; NID:9466582; PIDN:AB18585.1; PID:g1657692  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;  
Riley, M.; Colliado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,  
J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: B65161  
A:Molecule type: DNA  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Residues: 1-339 <BLAT>  
A:Cross-references: GB:AB000439; GB:U00096; NID:g1790036; PIDN:AAC76632.1;  
PID:g1790037; UMGP:B3608  
A:Experimental source: strain K-12, substrain MG1655  
A:Gene: gpsA  
C:Genetics:  
C:Superfamily: glycerol-3-phosphate dehydrogenase  
C:Keywords: oxidoreductase  
S47829 Length: 339 February 11, 2000 15:51 Type: P Check: 4443 ..

1 MNQRNASMTV IGAGSYGTAL AITLARNGHE VLMGHDPHE IATLERDRCN  
51 AAFLEDPVPP DTLHLESDLA TALAASRNIL VVPSHVFFE VLKQIKPLMR  
101 PPARLVWARK GLEAETGRLL QDVAREALGD QIPLAVISGP TFARKEIAGL  
151 PRAISLASTD QTEADDIQOL LHCGKSFERY SNPDIFGVL GGAVKRVIAI  
201 GAGMSDGIKF GANARITALIT RGLAEXSRIG AALGADPAIF MGWAGIGDLV  
251 LCTDNQSRN RRGGMILGOG MVOVSAQEKI GÖVVEGYRNT KEVELAHRE  
301 GVEMPTTEBI YÖVLYCGKNA REPAALTLGR ARKDERSSH  
301 GVEMPTTEBI YÖVLYCGKNA REPAALTLGR ARKDERSSH

!!AA\_SEQUENCE 1.0  
P1:A44132 - D-aspartate oxidase (EC 1.4.3.1) - bovine

C:Species: Bos primigenius taurus (cattle)  
C>Date: 27-Apr-1993 #sequence\_revision 18-Nov-1994 #text\_change 21-Aug-1998  
C:Accession: A44132  
R:Negri, A.; Cegallan, F.; Tedeschi, G.; Simonic, T.; Ronchi, S.  
J. Biol. Chem. 267, 11865-11871, 1992  
A>Title: The primary structure of the flavoprotein D-aspartate oxidase from beef kidney.  
A:Reference number: A44132; MUID:92291057  
A:Accession: A44132  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-338 <NEG>  
A:Experimental source: kidney  
A>Note: sequence extracted from NCBI backbone (NCBIP:106605)  
C:Superfamily: D-amino-acid oxidase  
C:Keywords: oxidoreductase

A44132 Length: 338 February 11, 2000 15:51 Type: P Check: 6742 ..

1 MDTVRIAVVG AGVGLSTAV CISKWPGCS ITVIDSKFTPT ETTSDVAAGM  
51 LIPTTPDTP IQKQKWFKE TPDHLEAIIV SAEADAGVI LVSGWQIFQS  
101 IPTEVPYPA DVLGFRKMT KDEKKFPQH VGHAFITLK CEGPVIPLW  
151 QKRVKNGGL ILTRIEDLM ELHPSFDIV NCSGLSROL AGDSKIFPV  
201 GOVLKQVAP VKHFIRDSG LTIYIPGVN VILGTRQKG DMNLSPDAEI  
251 SKELSRCA LEPSLRGAYD LREVRGLRPT RGVLEKEL LAQDSRLPV  
301 VHHYHGSGG IAHMWTALE ATRLVNECVQ VLRTPAPK  
!!AA\_SEQUENCE 1.0  
P1:JC5438 - D-aspartate oxidase (EC 1.4.3.1) type III 1 - human  
N:Alternate names: aspartic oxidase  
C:Species: Homo sapiens (man)  
C>Date: 17-Jun-1997 #sequence\_revision 18-Jul-1997 #text\_change 11-Jun-1999  
C:Accession: JC5438  
R:Seoyama, C.; Miura, R.  
J. Biochem. 121, 798-803, 1997  
A>Title: Structural and functional characterization of the human brain D-aspartate oxidase.  
A:Reference number: JC5438; MUID:97306065  
A:Accession: JC5438  
A:Molecule type: mRNA  
A:Residues: 1-341 <SET>  
A:Cross-references: DDBJ:D89858; NID:91742023; PIDN:BAA14031.1; PID:d1014731; PID:91742024  
A:Experimental source: brain  
A>Note: the authors translated the codon GAA for residue 83 as Asp and TGG for residue 95 as Leu  
C:Comment: This enzyme is a flavoprotein that catalyzes the oxidative deamination of dicarboxylic D-amino acids to give the corresponding alpha-keto acids and ammonia. It is involved in the regulation of the central nervous system and/or the development of the brain.  
C:Superfamily: D-amino-acid oxidase  
C:Keywords: oxidoreductase  
F:223/Active site: Tyr #status predicted  
F:278/Active site: Arg #status predicted

JC5438 Length: 341 February 11, 2000 15:51 Type: P Check: 7167 ..

1 MDTVRIAVVG AGVGLSTAV CISKWPGCS ITVIDSKFTPT ETTSDVAAGM  
51 LIPTTPDTP IQKQKWFKE TPDHLEAIIV SAEADAGVI LVSGWQIFQS  
101 IPTEVPYPA DVLGFRKMT KDEKKFPQH VGHAFITLK CEGPVIPLW  
151 QKRVKNGGL ILTRIEDLM ELHPSFDIV NCSGLSROL AGDSKIFPV  
201 GOVLKQVAP VKHFIRDSG LTIYIPGVN VILGTRQKG DMNLSPDAEI

251 SREILSRCA LEPSLRGACN IREKVGRLPY RGVRLQTEL LARDGRLPV  
301 VHHYHGSGG ISVHWGTALE AARLVSECVH ALRTPIPKSN L

!!AA\_SEQUENCE 1.0  
P1:JC5439 - D-aspartate oxidase (EC 1.4.3.1) type III 2 - human  
N:Alternate names: aspartic oxidase  
C:Species: Homo sapiens (man)  
C>Date: 17-Jun-1997 #sequence\_revision 18-Jul-1997 #text\_change 07-May-1999  
C:Accession: JC5439  
R:Seoyama, C.; Miura, R.  
J. Biochem. 121, 798-803, 1997  
A>Title: Structural and functional characterization of the human brain D-aspartate oxidase.  
A:Reference number: JC5438; MUID:97306065  
A:Accession: JC5439  
A:Molecule type: mRNA  
A:Residues: 1-282 <SET>  
A:Cross-references: DDBJ:D89858  
A:Experimental source: brain  
A>Note: the authors translated the codon GAA for residue 83 as Asp  
C:Comment: This enzyme is a flavoprotein that catalyzes the oxidative deamination of dicarboxylic D-amino acids to give the corresponding alpha-keto acids and ammonia. It is involved in the regulation of the central nervous system and/or the development of the brain.  
C:Superfamily: D-amino-acid oxidase  
C:Keywords: oxidoreductase

JC5439 Length: 282 February 11, 2000 15:51 Type: P Check: 934 ..

1 MDTVRIAVVG AGVGLSTAV CISKWPGCS ITVIDSKFTPT ETTSDVAAGM  
51 LIPTTPDTP IQKQKWFKE TPDHLEAIIV SAEADAGVI LVSGIKSGG  
101 WTLTRIEDL WEHPSFDIV VNCGLGSRQ LAGDSKIFPV RGQVLOVAP  
151 WVEHFRDGS GLTIYIPGTS HTVLGTRQK GDMNLSPDAE NSREILSRCC  
201 ALEPSLHGAC NIREKVLNP YRPGVRLQTE LARDGRLPV VHHYHGSGG  
251 GISVHWGTAL EARLVSECV HALRTPIPKS NL

!!AA\_SEQUENCE 1.0  
P1:S25422 - NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - Japanese quail mitochondrion (SCC1) (Fragment)  
C:Species: mitochondrion Coturnix coturnix japonica (Japanese quail)  
C>Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 20-Aug-1999  
C:Accession: S25422  
R:Desjardins, P.; Morais, R.  
J. Mol. Evol. 32, 153-161, 1991  
A>Title: Nucleotide sequence and evolution of coding and noncoding regions of a quail mitochondrial genome.  
A:Reference number: S25422; MUID:91178819  
A:Accession: S25422  
A:Molecule type: DNA  
A:Residues: 1-72 <DBS>  
A:Cross-references: EMBL:X57246; NID:912849; PIDN:CAA40522.1; PID:912850  
C:Gene: NDI  
A:Genome: mitochondrion  
A:Genetic code: SCC1  
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 1  
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation; oxidoreductase; respiratory chain

S25422 Length: 72 February 11, 2000 15:51 Type: P Check: 6471 ..

1 SFLSPSELFL SITLAKVLL IJSSFLWIRA SYPRFRYDQL MHLMLKNFLP  
51 LTIAMCLMHT SMPISYAGLP PA

!!AA\_SEQUENCE 1.0  
P1:S08622 - hydrogenase (EC 1.18.99.1) 3 chain 4 - Escherichia coli

N:Alternate names: formate hydrogenlyase chain 4; hydrogenase-3 protein D  
C:Species: *Escherichia coli*  
C>Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 26-Aug-1999  
C:Accession: S08622; F65052  
R:Boehm, R.; Sauter, M.; Boeck, A.  
Mol. Microbiol. 4, 231-243, 1990  
A:Title: Nucleotide sequence and expression of an operon in *Escherichia coli* coding for formate hydrogenlyase components.  
A:Reference number: S08619; MUID:90251163  
A:Accession: S08622  
A:Molecule type: DNA  
A:Residues: 1-307 <BOE>  
A:Cross-references: EMBL:X17506; NID:9556890; PID:941683  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of *Escherichia coli* K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: F65052  
A>Status: Preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-307 <BLAT>  
A:Cross-references: GB:AE000356; GB:U00096; NID:92367153; PID:91789077; UMG:b2722  
A:Experimental source: strain K-12, substrain MG1655  
C:Genetics:  
A:Gene: hycD  
A:Map position: 58-59 min  
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 1  
C:Keywords: oxidoreductase; transmembrane protein

S08622 Length: 307 February 11, 2000 15:51 Type: P Check: 675

1 MSVLPILQIA LVFAVAPLL SGITRVARAR LHNRRGPGVL QERYDIRLL  
51 GROSQGPAS GWFRLLTPY MYGVALTIAT ALPVTYVGP LPQLDLYTL  
101 LYLFAIAFF FAISGLDTGS PETAIGASRE AMGLVLEPM LILGLWVAQ  
151 VAGSTNISNI TDTVHMPLS GSIPVLALC ACAFAPIEM GLRFDLAEV  
201 EQELQEGPLS EYSGSGGVW KMGISLKLVL VLOMFVGVFI PMGQMETFLA  
251 GGLLALVIA IVKLVGVLV IALFNSMAR LRDIITPRIT WAGFGFATLA  
301 FVSLIAA

11AA:SEQUENCE 1.0  
P1:T11338 - NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 - aardvark  
C:Species: *mitochondrion Oryzteropus afer* (aardvark)  
C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 13-Aug-1999  
C:Accession: T11338  
R:Arnason, U.; Gullberg, A.; Janke, A.  
Proc. R. Soc. Lond. B Biol. Sci. 266, 339-345, 1999  
A:Title: The mitochondrial DNA molecule of the aardvark, *Oryzteropus afer*, and the position of the tubulin gene in the eutherian tree.  
A:Reference number: Z17633; MUID:99197468  
A:Accession: T11338  
A>Status: Preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-347 <ARN>  
A:Cross-references: EMBL:X18475; NID:e1427536; PID:e1427538; PIDN:CAB41622.1  
C:Genetics:  
A:Genome: mitochondrion  
A:Genetic code: SGC1  
A>Note: NADH2  
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 2  
C:Keywords: mitochondrion; NAD; oxidoreductase

T11338 Length: 347 February 11, 2000 15:51 Type: P Check: 2220

1 NMPLFLMIM MTLSCGIIIT MLSSHWFTAW MGLEMMNFAL IPIIKNNHP  
51 RSEIATATF LTOATASMLL MIALNLNL SGMSTISIP SSFTSILVT  
101 ALIMKLGAP EHWLPEVTO GTHSSGLLI LTMOKLAPS ILYOISNIN  
151 PTLMTSAFL SILIGNGGI NOTQRLKLA YSSIAHGMW ANILYVPTL  
201 TLNLILYIT LTSAFMLI INYSTSTSL SLIMNKTPLI SLAMAILLS  
251 MGELPLSGF IPKWLMOEL TKNLNAIPE LMATLALNL YFYMILYSS  
301 SLTIIPSMNN MKIKMFNNH NIYPIISPII LSTMLPLT PYMSLF

11AA:SEQUENCE 1.0  
P1:T11159 - NADH dehydrogenase 3 - hardbacked tick (*Rhipicephalus sanguineus*) mitochondrion  
C:Species: *mitochondrion Rhipicephalus sanguineus*  
C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 20-Sep-1999  
C:Accession: T11159  
R:Black IV, W.C.; Roehrdaiz, R.L.  
Mol. Biol. Evol. 15, 1772-1785, 1998  
A:Title: Mitochondrial gene order is not conserved in arthropods: prostrate and metastriate tick mitochondrial genomes.  
A:Reference number: Z17252; MUID:99083443  
A:Accession: T11159  
A>Status: Preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-114 <BLA>  
A:Cross-references: EMBL:AF081829; NID:94164556; PID:94164563; PIDN:AAD05523.1  
C:Genetics:  
A:Genome: mitochondrion  
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 3  
C:Keywords: mitochondrion

T11159 Length: 114 February 11, 2000 15:51 Type: P Check: 979

1 MIFYLHITF LVCLIMLFL FSLGFGKKA KEKNSPEGC FDFSLSRVP  
51 FSLKEFEVGI VFLIFDEIV VILPPLVMK TKNLMEVFSF TELNPLIVIG  
101 LLYEFKYSML DRLK

11AA:SEQUENCE 1.0  
F1:S42244 - NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Anser caerulescens mitochondrion (SGC1) (fragment)  
C:Species: *mitochondrion Anser caerulescens*  
C>Date: 06-Dec-1996 #sequence\_revision 06-Dec-1996 #text\_change 11-Jun-1999  
C:Accession: S42244  
R:Quinn, T.W.; Wilson, A.C.  
J. Mol. Biol. 37, 417-425, 1993  
A:Title: Sequence evolution in and around the mitochondrial control region in birds.  
A:Reference number: S42243; MUID:94141942  
A:Accession: S42244  
A:Molecule type: DNA  
A:Residues: 1-214 <QU>  
A:Cross-references: EMBL:X75772; NID:9450491; PIDN:CAA53393.1; PID:91164899  
C:Genetics:  
A:Gene: ND5  
A:Genome: mitochondrion  
A:Genetic code: SGC1  
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5  
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation; oxidoreductase; respiratory chain

S42244 Length: 214 February 11, 2000 15:51 Type: P Check: 1655

1 GSVDLQPSLN TSYLNTALL LTLMTAFYA TYSIMRTLV QAGOTRIPP  
51 VSMNNENPLI TAPLRLALG SITAGMIIS FITPKTPPM TMDLITRTAA  
101 ILMTITGIL ALSLNMTHI LTYKPNPLM NSSALGYFN PLVHREFSKN

151 LKERGNIAL HLIDSLWLKK MGPEGLAELQ VAASKATSM HTGLIKAYLG  
201 SFALSLIVMI LMTM

IIAA\_SEQUENCE 1.0  
P1:F64999 - NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) I chain J -  
Escherichia coli  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 20-Aug-1999  
C:Accession: F64999; S38310; S37067  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;  
Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,  
J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: F64999  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-184 <BLAT>  
A:Cross-references: GB:AEO00317; GB:U00096; NID:91788605; PIDN:AMC75340.1;  
PID:91788616; UMG:P.b2280  
A:Experimental source: strain K-12, substrain MG1655  
R:Weidner, U.; Geier, S.; Plock, A.; Friedrich, T.; Lell, H.; Welts, H.  
J. Mol. Biol. 233, 109-122, 1993  
A:Title: The gene locus of the proton-translocating NADH:ubiquinone  
oxidoreductase in Escherichia coli. Organization of the 14 genes and  
relationship between the derived proteins and subunits of mitochondrial complex  
I  
A:Reference number: S38310; MUID:93389724  
A:Accession: S38310  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-84; 'T', 86-125; 'R', 127-184 <ME2>  
A:Cross-references: EMBL:X68301; NID:9444012; PIDN:CAA6369.1; PID:9397907  
A:Experimental source: strain AN387  
C:Genetics:  
A:Gene: nuoJ  
A:Map position: 49.5 min  
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 6  
C:Keywords: membrane-associated complex; NAD; oxidoreductase; transmembrane  
protein  
F64999 Length: 184 February 11, 2000 15:51 Type: P Check: 2909 ..

1 MEFAFYICGL IALIALRLVI THTNPVHAL YLITSLAIS GVEFSLGAYF  
51 AGALEIIVYA GAIMVLFYRV VMMLNGSSE IEEERQWLKP QVMIGPAIIS  
101 AIMLVIVYA ILGVNDQGD GTPISAKAVG ITLFGPVYLA VELASMLLIA  
151 GLVYAFHVGR EERAGEVLN RKDSAKAKT EEHA

IIAA\_SEQUENCE 1.0  
P1:S52970 - NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 6 - honeybee  
mitochondrion (SGC4)  
C:Species: mitochondrion Apis mellifera (honeybee)  
C:Date: 15-Jul-1995 #sequence\_revision 10-Nov-1995 #text\_change 20-Aug-1999  
C:Accession: S52970  
R:Crozler, R.H.; Crozier, Y.C.  
Genetics 133, 97-117, 1993  
A:Title: The mitochondrial genome of the honeybee Apis mellifera: complete  
sequence and genome organization.  
A:Reference number: S52960; MUID:93114603  
A:Accession: S52970  
A:Molecule type: DNA  
A:Residues: 1-167 <CRO>  
A:Cross-references: EMBL:L06178; NID:9336279; PIDN:AAB96808.1; PID:9829009  
C:Genetics:  
A:Genome: mitochondrion  
A:Genetic code: SGC4

IIAA\_SEQUENCE 1.0  
P1:S78138 - cytochrome-c oxidase (EC 1.9.3.1) chain III - Reclinomonas  
americana (ATCC 50394) mitochondrion  
C:Species: mitochondrion Reclinomonas americana  
A:Variety: ATCC 50394  
C:Date: 29-Jan-1998 #sequence\_revision 06-Feb-1998 #text\_change 20-Aug-1999  
C:Accession: S78138  
R:Lang, B.F.; Burger, G.; O'Kelly, C.J.; Cedergren, R.; Golding, G.B.; Lemieux,  
C.; Sankoff, D.; Turmel, M.; Gray, M.W.  
Nature 387, 493-497, 1997  
A:Title: An ancestral mitochondrial DNA resembling a eubacterial genome in  
miniature.  
A:Reference number: S78127; MUID:97311393  
A:Accession: S78138  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-267 <LAN>  
A:Cross-references: EMBL:AF007261; NID:92258325; PIDN:AD11871.1; PID:92258337  
A:Experimental source: ATCC 50394  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June  
1997  
C:Genetics:  
A:Gene: cox3  
A:Genome: mitochondrion  
C:Superfamily: cytochrome-c oxidase chain III  
C:Keywords: electron transfer; membrane-associated complex; mitochondrial inner  
membrane; mitochondrion; oxidative phosphorylation; oxidoreductase; respiratory  
chain; transmembrane protein  
F:4-18/Domain: mitochondrial matrix #status predicted <MM1>  
F:19-37/Domain: transmembrane #status predicted <TM01>  
F:38-45/Domain: intracristal #status predicted <ITC1>  
F:46-71/Domain: transmembrane #status predicted <TM02>  
F:72-77/Domain: mitochondrial matrix #status predicted <MM2>  
F:78-110/Domain: transmembrane #status predicted <TM03>  
F:111-133/Domain: intracristal #status predicted <ITC2>  
F:134-157/Domain: transmembrane #status predicted <TM04>  
F:158-160/Domain: mitochondrial matrix #status predicted <MM3>  
F:161-188/Domain: transmembrane #status predicted <TM05>  
F:189-195/Domain: intracristal #status predicted <ITC3>  
F:196-228/Domain: transmembrane #status predicted <TM06>  
F:229-237/Domain: mitochondrial matrix #status predicted <MM4>  
F:238-261/Domain: transmembrane #status predicted <TM07>  
F:262-266/Domain: intracristal #status predicted <ITC4>

S78138 Length: 267 February 11, 2000 15:51 Type: P Check: 8115 ..

1 MSQTFVKRHP YATVDSPWP LITSIGTICS TFGGVWFHS YPNGFIAL  
51 GVSTILFSLY AMCRDIYREG TYOGHTAAV QNGLRIGMIL FISEVMFV  
101 SFMAFFHSS LSPITIEIGAV WPQCIETLN AMDVPFNTV IILMSGATV  
151 WSHAMINGN RTQSIIGLIF TIIAVFTFG LQVNEYREAS FSIADGIYS  
201 TFYATGFGH FHVIVGCMU SYCLVRELY HFTTTHHGF EASAWMHHV  
251 DVWMLFETI IYWGNGN

```

!!AA_SEQUENCE 1.0
P1:S07557 - cytochrome-c oxidase (EC 1.9.3.1) chain III - bracket fungus
(Schizophyllum commune) mitochondrion
C1:Species: mitochondrion Schizophyllum commune
C1:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 20-Aug-1999
C1:Accession: S07557
R1:Phelps, L.G.; Burke, J.M.; Ullrich, R.C.; Novotny, C.P.
Curr. Genet. 14, 401-403, 1988
A1:Title: Nucleotide base sequence of the mitochondrial COIII gene of
Schizophyllum commune
A1:Reference number: S07557; MUID:89106272
A1:Accession: S07557
A1:Molecule type: DNA
A1:Residues: 1268 <PHE>
A1:Cross-references: EMBL:X15748
C1:Genetics:
A1:Gene: COIII
A1:Genome: mitochondrion
C1:Superfamily: cytochrome-c oxidase chain III
C1:Keywords: electron transfer; membrane-associated complex; mitochondrial inner
membrane; mitochondrion; oxidative phosphorylation; oxidoreductase; respiratory
chain; transmembrane protein

S07557 Length: 268 February 11, 2000 15:51 Type: P Check: 7659 ..

1 MTSIKYQSF SAHLVQSPW PILVFSILFN LAIGVLTWH GYSHSTTFD
51 LGLAVTQSI LMTVDIVIE GSFGLDHTKQ VQEGILIGFI LFIISVFAF
101 ISVMAFVHS ALSPAVELGS TWPVGIIP LDFESLPLFNT IILSSGAFV
151 TYGHHAFVSG KRDSIIGLF LTVALALIFS YFOAFEYIHA GFSMSDVFV
201 TVFASVSGIH GIHVMIGTLF LFVSGFLQVN YQTKKHNGJ LETSLVWHF
251 VDLWMLFEL VYFVWGCA

!!AA_SEQUENCE 1.0
P1:T09997 - lipoygenase (EC 1.13.11.12) - southern Asian dodder (fragment)
C1:Species: Cuscuta reflexa (southern Asian dodder)
C1:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C1:Accession: T09997
R1:Subramaniam, K.; Mahadevan, S.
Submitted to the EMBL Data Library, January 1994
A1:Description: Cloning and characterization of a cDNA coding for lipoygenase
associated with cytokinin-induced haustoria formation in Cuscuta reflexa.
A1:Reference number: Z16913
A1:Accession: T09997
A1>Status: preliminary; translated from GB/EMBL/DBJ
A1:Molecule type: mRNA
A1:Residues: 1385 <SUB>
A1:Cross-references: EMBL:U05041; NID:9450230; PID:9450231
C1:Genetics:
A1:Gene: lxx
A1:Function:
A1:Description: hydroperoxidation of unsaturated fatty acids
A1:Superfamily: lipoygenase
C1:Keywords: fatty acid oxidation; oxidoreductase

T09997 Length: 385 February 11, 2000 15:51 Type: P Check: 1856 ..

1 WMESHSGSKC FAIPGMPLVA WMKTKAKAHV LAHDSGYHOL VSHMLRTCA
51 IEPYIIASNR QLSMLPTIFR LKPHFRITM EINNLAERAL INANVIST
101 FSPGKSHMLT SSMAVDEMG FDLQALPADL ISRGIAEDP TAPHGLKALAI
151 EEPYPRANDL VMDIIEKEMV TDYVTHYPE SSRVSESTEL QEMWKEIRTV
201 GHADKRDAGG WPLDKTPEDL IGIITITVY CSGHNAVNF GQYVDAQIFP
251 NRPITARTKM PTHDPSESEW EHPMKKPEDT LKCFPSQLQ ARIKVAIVDV

!!AA_SEQUENCE 1.0
P1:S43771 - phosphatidylcholine desaturase (EC 1.3.1.35) - Synecchococcus sp.
(strain PCC 7002)
N1:Alternate names: Delta-12 desaturase; fatty acid desaturase (EC 1.14.99.-)
(misidentification)
C1:Species: Synecchococcus sp.
A1:Variety: PCC 7002
C1:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 28-May-1999
C1:Accession: S43771
R1:Sakamoto, T.; Wada, H.; Nishida, I.; Ohmori, M.; Murata, N.
Plant Mol Biol. 24, 643-650, 1994
A1:Title: Identification of conserved domains in the Delta-12 desaturases of
cyanobacteria
A1:Reference number: S43770; MUID:94207189
A1:Accession: S43771
A1>Status: preliminary
A1:Molecule type: DNA
A1:Residues: 1347 <SAK>
A1:Cross-references: GB:D13779; NID:9488510; PIDN:BA002922.1; PID:dl003428;
PID:9488511
C1:Superfamily: omega-3 fatty acid desaturase
C1:Keywords: oxidoreductase

S43771 Length: 347 February 11, 2000 15:51 Type: P Check: 2848 ..

1 MTSVTRPSA TLLEKHPNL RLRLDILTL RSYVEINPLK AMSVLLSVA
51 AVVGCYALLA IAPWYLLPV WELTGTLLTG FVYIGHDGH RSFRKKNVN
101 NLVGHIAFLP LIYFPHSMRI LNNHHRXTN NMDENANAP FTPELYDSDP
151 AFKAVYRAI RGLIMWLASV IHOULKHEW FAFEGKQROD VRFSALEVLI
201 AGAIFPVNF YTLGVWGVK FWLMPWLGX FMSFTFLVH HTVEIDPFSY
251 RDKWEALAQ LSGTVACDYP KWVEYLCHDI NVHVPHLST GIPSYNLKKA
301 VASIKQNGE YIVETKFSWE LKRAITEGCH LYDAENHYIS FAQHOKR

!!AA_SEQUENCE 1.0
P1:S74692 - tRNA(m137)methyltransferase - Synecchocystis sp. (strain PCC 6803)
N1:Alternate names: protein s111198
C1:Species: Synecchocystis sp.
A1:Variety: PCC 6803
C1:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 18-Jun-1999
C1:Accession: S74692
R1:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.;
Miyajima, N.; Hirosewa, M.; Sugiyama, M.; Sasamoto, S.; Kimura, T.; Hosouchi,
T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naito, K.; Okumura, S.; Shimo, S.;
Takenuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 3, 109-136, 1996
A1:Title: Sequence analysis of the genome of the unicellular cyanobacterium
Synecchocystis sp. PCC6803. II. Sequence determination of the entire genome and
assignment of potential protein-coding regions.
A1:Reference number: S74322; MUID:97061201
A1:Accession: S74692
A1>Status: nucleic acid sequence not shown; translation not shown
A1:Molecule type: DNA
A1:Residues: 1231 <KAN>
A1:Cross-references: EMBL:D90901; GB:AB001339; NID:91651897; PIDN:BA16843.1;
PID:dl017576; PID:91651917
A1>Note: the nucleotide sequence was submitted to the EMBL Data Library, June
1996
C1:Genetics:
A1:Gene: trmd
A1:Start codon: GTG
C1:Superfamily: tRNA (guanine-N1) methyltransferase
C1:Keywords: methyltransferase

```

S74692 Length: 231 February 11, 2000 15:51 Type: P Check: 6210 ..

1 MOPDVLTLP DFTSPLOSG LLGKALEKAI ASYNLINPRD FTTPKHRRVD  
 51 DEPYGGGVGM VIKPEPIFAA VESIPVLRSR EVILMPOQ PMDQALFREL  
 101 TNYDQVLIC GHYEGDERV COLTVREVSL GDFVLTCGEI PALTLINGVI  
 151 RLPGTVGKE ASLIAESFST DLLDYPHYTR PPFVRLAIVP PVLISGNHQA  
 201 IAQRLEQOE ERTQGRRPDL WQKWQDRQPS P

!!AA-SEQUENCE 1.0  
 P1:A71119 - probable aspartate carbamoyltransferase catalytic chain -  
 C:Species: Pyrococcus horikoshii  
 C>Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 18-Jun-1999  
 C:Accession: A71119  
 R:Kawabuchi, Y.; Sawada, M.; Horikawa, H.; Hatake, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohkuni, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kishida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.  
 DNA Res. 5, 55-76, 1998  
 A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.  
 A:Reference number: A71000; MUID:98344137  
 A:Accession: A71119  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-308 <RAW>  
 A:Cross-references: GB:AP000003; NID:93236130; PIDN:BA29811.1; PID:dl030754; PID:93257128  
 A:Experimental source: strain OT3  
 A>Note: this accession replaces an interim accession for a sequence replaced by GenBank

C:Gene: PH0720  
 C:Superfamily: ornithine carbamoyltransferase: aspartate/ornithine carbamoyltransferase homology  
 F:6-305/Domain: aspartate/ornithine carbamoyltransferase homology <ACT>  
 A71119 Length: 308 February 11, 2000 15:51 Type: P Check: 9980 ..

1 MEMKGRDVIS IDFSKEDIE VILSTARLE KEMKEKGQLE YAKGKIATL  
 51 FEPSPTRTRL SPESAMHRLG GSVIGFAEAS TSSVKKGESL RDTIKIVEQY  
 101 SDVIYIRHPK EGARLALEV ADIPVINAGD GSNQHTQTL LDLYTIKKEF  
 151 GTIDGLKIGL LGDLKYGRTV HSLAEALAFY DVELYIISPE LLRMPKHIVE  
 201 ELNRGKRTV ETTKLEEVIG ELVDLYVTRI QKERFDEDE YLAKXGSGYQ  
 251 NKILLENVAD SLRIMHPLPR VDEIHPEYDK TKHAIYFKQV FNGVPVPMAL  
 301 LALVGVY

!!AA-SEQUENCE 1.0  
 P1:A70959 - probable aspartate carbamoyltransferase - Mycobacterium tuberculosis (strain H37Rv)  
 C:Species: Mycobacterium tuberculosis  
 C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 26-Aug-1999  
 C:Accession: A70959  
 R:Colet, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekale, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: A70959  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-319 <COI>  
 A:Cross-references: GB:281011; GB:AL123456; NID:93242274; PID:e275151; PID:91621262  
 A:Experimental source: strain H37Rv  
 C:Species: PyTB  
 C:Superfamily: ornithine carbamoyltransferase: aspartate/ornithine carbamoyltransferase homology  
 F:4-307/Domain: aspartate/ornithine carbamoyltransferase homology <ACT>  
 A70959 Length: 319 February 11, 2000 15:51 Type: P Check: 1547 ..

1 MPRHLITRA DLSDRDATAI LDDADRFPAQ LVGRDIRKLP TLGRRTVYM  
 51 EYENSTRIRV SEVAGKMS ADVINVSAG SVYKGESLR DIALTLRAAG  
 101 ADALITRHPA SGAHLLAOW TGAHNDGPV INAGDGTHER PTQALLDALP  
 151 IRORLGIGIG RRIIVGDI LSHVARSNVM LIDTLGAEVV LVAPPTLLPV  
 201 GVTGWPAIVS HDPAELPAA DAVIMLRVQA ERMNGGFPS VREYSVRYGL  
 251 TERQAMLPG HAVYLHPQPM VRCMEITSSV ADSSQSAVLO QVSNQGVYRM  
 301 AVLFHVLVGA QDAKKEGAA

!!AA-SEQUENCE 1.0  
 P1:C70815 - probable beta-ketoadipyl CoA thiolase - Mycobacterium tuberculosis (strain H37Rv)  
 C:Species: Mycobacterium tuberculosis  
 C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 26-Aug-1999  
 C:Accession: C70815  
 R:Colet, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekale, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: C70815  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-403 <COI>  
 A:Cross-references: GB:AL022004; GB:AL123456; NID:93261550; PID:el254001; PID:92916918  
 A:Experimental source: strain H37Rv  
 C:Species: Tada  
 C:Superfamily: acetyl-CoA acetyltransferase  
 C70815 Length: 403 February 11, 2000 15:51 Type: P Check: 556 ..

1 MSEAFIYDA IRTPRGOKN GSLHEVPPS LVVGLIDELR KRHPDDENL  
 51 ISDVIIGCVS PVGQGGDIA RAAYLVASGMP VTSGGVQLNR FCASGLEAVN  
 101 TAAQXKRSWM DDVILAGSVE SMSRVMGSD GGAGLDPAT NYDVMFVPOS  
 151 IGDGLTATIE GFSREDPAY ALRSQQAEE AMSGTYFAKS VYVVRQNGL  
 201 LITDDEHNR PDITKEGLK LKPAFEGIAA LGFDDVALQ KYHWKXINH



```

251 VHTGNSSGI VDGALVIG SAAGKLOGL TPRARIYATA TSGADPVIWL
301 TGPTPATKRV LDRAGLTVDL IDLFELNEAF ASVYLFQKQD LNIPEKLVV
351 NGGALAMGHP LGATGAMIIG TWDLELERN ARRALTITICI GGGMGVATII
401 ERV

!!AA_SEQUENCE 1.0
P1:12981 glucuronosyl transferase-like protein - Arabidopsis thaliana
N:Alternate names: protein T6H20.280
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 20-Sep-1999
C:Accession: 12981
R:Choliste, N.; Robert, C.; Bottier, P.; Wincker, P.; Cattoi, L.;
Artiguenave, F.; Saulin, W.; Weissenbach, J.; Mevius, H.W.; Mayer, K.F.X.;
Lemcke, K.; Schueller, C.; Queller, F.; Salanoubat, M.
Submitted to the Protein Sequence Database, July 1999
A:Reference number: 217586
A:Accession: 12981
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-452 <CHO>
A:Cross-references: EMBL:AL096859; GSPDB:GN00061; ATSP:T6H20.280
A:Experimental source: cultivar Columbia; BAC clone T6H20
C:Genetics:
A:Gene: ATSP:T6H20.280
A:Map position: 3
A:Introns: 161/1
C:Superfamily: flavonol O3-glucosyltransferase

T12981 Length: 452 February 11, 2000 15:51 Type: P Check: 8558

1 MEKRVKRII VLVPAAGH VTPMOLGKA LQSKGLITV AQRFQIIGS
51 SIQHFGDFP VTIPESLPQS ESKKLPAEV LMNNTKSEA SFPECISQIS
101 MOGNDIACI IYDKIMFCE AAKKEKIPS VIFSTSSATI QVCYCVLSEL
151 SAEKFLDMK DPEKDKYLE GLHPLRYKDL PISGPGLEP LLMKREYVN
201 KRASAVIIN TASCLESISL SWIQOELGIP VYPLGLHIT ASSPGSLDL
251 EDMSCIEWIN KOKPRSVYI SLGKAMET KEMEMANGL LNSNOPILVY
301 IRPGSVAGEF WIELPEEVI KMYTERGYIA KPAQIEVLG HPVAGGFWSH
351 CGNNTLESI VEGVPMICRP LQGEOKINAM YIESYWKIGI QLEGEVEREG
401 VERAVKRLII DEGAAMRER ALDLKELNA SVRSGSSYN ALDELVKFLN
451 TE

!!AA_SEQUENCE 1.0
P1:140986 streptomycin/spectinomycin adenylyltransferase - Escherichia coli
C:Species: Escherichia coli
C>Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 18-Jun-1999
C:Accession: 140986
R:Hollingshead, S.; Vapnek, D.
Plasmid 13, 17-30, 1985
A:Title: Nucleotide sequence analysis of a gene encoding a
streptomycin/spectinomycin adenylyltransferase.
A:Reference number: 140986; MUID:85191077
C:Accession: 140986
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-323 <RES>
A:Cross-references: EMBL:X02340; NID:940880; PIDN:CAA26199.1; PID:9581035
C:Superfamily: streptomycin 3'-adenylyltransferase

I40986 Length: 323 February 11, 2000 15:51 Type: P Check: 7621

1 MTAQNRFSM LVMTCFFGV OSNPRASKOO ARVAVGCLM LMSNDVTOQ
51 GSRPKTKINI MREAVIAEVS TOLSEYVGI ERHLEPTILA VHYGSANGD
101 GLKPHSIDL LVTYVRIDE TTRRALINDL LETSASGES EILRAVEYI
151 VVHDDIIPMR YPAKRELOFG EMORNDILAG IFEPATIDID LAILTKARE
201 HSAVLGPAA EELDPVEEQ DFEALNETL TLWNSPPDMA GDERNVLTIL
251 SRIVYSAVIG KIAPDVNAD WAMERLPAQY QPVILARQA YLGOEDRILA
301 SRADQLEERY HYKGEITKV VGR

!!AA_SEQUENCE 1.0
P1:165760 alcohol sulfoxyltransferase (EC 2.8.2.2) - rat
N:Alternate names: hydroxysteroid sulfoxyltransferase; senescence marker protein 2B
C:Species: Rattus norvegicus (Norway rat)
C>Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 18-Jun-1999
C:Accession: 165760; 152407
R:Matabe, T.; Ogura, K.; Satsukawa, M.; Okuda, H.; Hiratsuka, A.
Chem. Biol. Interact. 92, 87-105, 1994.
A:Title: Molecular cloning and functions of rat liver hydroxysteroid
sulfoxyltransferase catalysing covalent binding of carcinogenic polycyclic
arylimethanols to DNA.
A:Reference number: 152849; MUID:94306585
A:Accession: 165760
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-285 <RES>
A:Cross-references: GB:D14989; NID:9303801; PIDN:BA03634.1; PID:9303802
R:Song, C.S.; Kim, J.M.; Roy, A.K.; Chatterjee, B.
Biochemistry 29, 542-551, 1990
A:Title: Structure and regulation of the senescence marker 2 gene promoter.
A:Reference number: 152408; MUID:90148982
A:Accession: 152407
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 2-68 <RES2>
A:Cross-references: GB:M29302; NID:9206995; PIDN:AAA42152.1; PID:9206996
C:Genetics:
A:Introns: 46/1
C:Superfamily: alcohol sulfoxyltransferase
C:Keywords: sulfoxyltransferase

I65760 Length: 285 February 11, 2000 15:51 Type: P Check: 800

1 MMSDYTFEG IPPPAFWFSK EILENSCKF VKKEDDILIL TYRSGTNWL
51 IEIVCLIQTK GDPKWIQSNP IMDRSPWIEF GSGYDKLTKM EGPRLMTSHL
101 PMHLSKSLF SSRKAVIYI RNPRDVLVA YFFWSKIALE KKPDSLGTYV
151 EMFLKGNVAY GSWFEHINQW LSMREWDNFI VLYYEDMKD TWSIKKICD
201 FLGKLEPDE LNLVLYKSSF QVAKENNSN YSLMEKELIL TGFTFMKGT
251 TNDWKNHFTV AQAEAFDKVF QEKWAGPPPG MEPPWE

!!AA_SEQUENCE 1.0
P1:S52265 dual specificity phosphatase (EC 3.1.3.-) 1 - rat
N:Alternate names: protein-tyrosine-phosphatase CL100;
protein-tyrosine-phosphatase, nonreceptor type 10
C:Species: Rattus norvegicus (Norway rat)
C>Date: 07-May-1995 #sequence_revision 21-Jul-1995 #text_change 22-Jun-1999
C:Accession: S52265
R:Muda, M.; Schlegel, W.; Arkinstall, S.
Submitted to the EMBL Data Library, January 1995
A:Description: Pathways regulating CL100 gene expression in pituitary cells.
A:Reference number: S52265
A:Accession: S52265
A:Status: preliminary

```

A:Molecule type: mRNA  
A:Residues: 1-367 <MCD>  
A:Cross-references: EMBL:X84004; NID:g642264; PIDN:CA58828.1; PID:g642265  
C:Superfamily: dual specificity phosphoprotein phosphatase 1; VHL-type dual  
C:Specificity phosphoprotein phosphatase homology  
C:Keywords: phosphoprotein; phosphoric monoester hydrolase  
F:181-312/Domain: VHL-type dual specificity phosphoprotein phosphatase homology  
<VHL>  
F:258/Active site: Cys (phosphocysteine intermediate) #status predicted  
F:266/Binding site: substrate phosphate (Arg) #status predicted  
S52265 Length: 367 February 11, 2000 15:51 Type: P Check: 6443 ..

1 MMEVGIIDA GGLRALLER AQCLLDOR SFFAFMGHI VGSVNRFT  
51 IVRRKAKAM GLEHIVPNTL LKRLLAGV HAVVLDEBS ALDCAKRG  
101 TLALAGALC REARSTQVFF LGGYEAFSA SCPELCKOS TPXGLSLPLS  
151 TSVPDSASG CSSCSTPLVD QGGPEIILSF LYIGSAVHAS RDMDLALCI  
201 TALINVSANC PNHEGHYQY KSIPYEDNKH ADISSWNEA IDFIDSINDA  
251 GGRVFNHCOA GISRSATIL AYLMRTNRVK LDEAFEFVKQ RRSIISPNS  
301 FMGQLAPES QVLAPHCSAE AGSPMAVLD RGTSTTVFN FVSIHVHT  
351 NSALNYIQSP ITTSPSC

!!AA\_SEQUENCE 1.0  
P1:TI2078 - ribonuclease (EC 3.1.-.-) - almond  
C:Species: Prunus dulcis (almond)  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 26-Aug-1999  
C:Accession: TI2078  
R:Ushijima, K.; Sassa, H.; Tao, R.; Yamane, H.; Dandekar, A.; Gradziel, T.; Hirano, H.  
Mol. Gen. Genet. 260, 261-268, 1998  
A:Title: Cloning and characterization of cDNAs of the S-RNases in almond (Prunus dulcis): primary structural features and sequence divergence of the rosaceae S-RNases.  
A:Reference number: 217403  
A:Accession: TI2078  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-231 <USH>  
A:Cross-references: EMBL:AB011469; NID:d1241928; PID:d1035644  
A:Experimental source: strain Mission, pistil  
C:Superfamily: Enterobacter ribonuclease  
C:Keywords: hydrolase  
TI2078 Length: 231 February 11, 2000 15:51 Type: P Check: 3061 ..

1 MAMKSSSAF LVAFAFLC FIMSGSIYV FQVQOMPPT NCVARKRC  
51 SNRPRLQFT IHGIMPSNIS NPTKSNCG SGNFTVSP KMRVAKRSM  
101 PDVESGNDTR FWEGMNKRG TCSESSLNOM QYERSHEW YSFNTEILK  
151 NASIVPHPTQ TWKYSIDVAP IKTAKRTPV LRCKRDPAN KSGPKTOLLH  
201 EYVCEYEA LKQIDCNRTA GCMNVDIRF Q

!!AA\_SEQUENCE 1.0  
P1:B71518 - probable lipoprotein signal peptidase - Chlamydia trachomatis (serotype D, strain UW3/Cx)  
C:Species: Chlamydia trachomatis  
C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 26-Aug-1999  
C:Accession: B71518  
R:Stephens, R.S.; Kaiman, S.; Lammel, C.J.; Fan, J.; Marthe, R.; Aravind, L.; Mitchell, W.S.; Oliniger, L.; Tatusov, R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.  
Science 282, 754-759, 1998  
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis.

A:Reference number: A71570; MUID:99000809  
A:Accession: B71518  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-167 <ARN>  
A:Cross-references: GB:AE001314; GB:AE001273; NID:g3328833; PID:g3328836  
A:Experimental source: serotype D, strain UW-3/Cx  
C:Genetics:  
A:Gene: lspa  
C:Superfamily: lipoprotein signal peptidase  
B71518 Length: 167 February 11, 2000 15:51 Type: P Check: 3917 ..

1 MPTSLPTFL TLLLASIDW VSKLVLLKS CQSPSSSAF LXYWGHFS  
51 FLIPSENEG AAFGLFTQYK IPLIFRCV ILGLALFLRI KXSILHRT  
101 VALTIIAGA LGNVGDILY GRVDFLSLS YSWRPFEN LADAFISIGT  
151 LLLIGHLYFT KESKYE

!!AA\_SEQUENCE 1.0  
P1:S26429 - probable dUTP pyrophosphatase (EC 3.6.1.23) - fowl adenovirus 1  
C:Species: Aviadenovirus gall (fowl adenovirus 1, CELO)  
C:Date: 07-May-1993 #sequence\_revision 02-Aug-1994 #text\_change 12-Apr-1995  
C:Accession: S26429  
R:Akopian, T.A.; Kaverina, E.N.; Naroditsky, B.S.; Tikhonenko, T.I.  
Submitted to the EMBL Data Library, October 1992  
A:Description: Avian adenovirus CELO: sequence of the DNA fragment (92-100%).  
A:Reference number: S26428  
A:Accession: S26429  
A:Molecule type: DNA  
A:Residues: 1-178 <AKO>  
A:Cross-references: EMBL:Z17216  
C:Superfamily: retroviral proteinase  
C:Keywords: hydrolase  
S26429 Length: 178 February 11, 2000 15:51 Type: P Check: 9885 ..

1 MDPFGSSVP PCSTDLPEP KLYVRUSPH APPVRRATG AAGYDLFSAY  
51 DIKVPARGRA LVPTDLVQF PGCGYRIAP RSGIAKFTI DVGACVIDPD  
101 YRGNVSVLF NSESEFNIR RGRVADQLI ERIWPELSE LTOIGETDRG  
151 ASGFGFTGMG AVDRNQRVL EWLTPGSR

!!AA\_SEQUENCE 1.0  
F1:S42592 - urease (EC 3.5.1.5), tissue-ubiquitous - soybean (fragment)  
C:Species: Glycine max (soybean)  
C:Date: 19-Mar-1997 #sequence\_revision 24-Jul-1997 #text\_change 07-May-1999  
C:Accession: S42592  
R:Totiskiy, R.S.; Griffin, J.D.; Yenofsky, R.L.; Polacco, J.C.  
Mol. Gen. Genet. 242, 404-414, 1994  
A:Title: A single gene (Edu) encodes the tissue-ubiquitous urease of soybean.  
A:Reference number: S42592; MUID:94166749  
A:Accession: S42592  
A:Molecule type: mRNA  
A:Residues: 1-131 <TOR>  
A:Cross-references: EMBL:S69179  
C:Genetics:  
A:Gene: Edu  
C:Superfamily: urease; urease 11k chain homology; urease 12k chain homology; urease 62k chain homology  
C:Keywords: hydrolase  
F:1-114/Domain: urease 62k chain homology (fragment) <U62>  
S42592 Length: 131 February 11, 2000 15:51 Type: P Check: 4170 ..

1 PSFEGAPEM YIKGVVANA DMGDPNASIP TPPEVKKRM FGLCKRAGA  
51 LSIAFVSKAA VDOVRHALTG LNKRVKAVN VRKLTLDK LNDSPQITV

```

101 DPNKVTATD GEVLTSPATT FVPLSRNYFL F
11AA_SEQUENCE 1.0
P1:151663 - arginase 1 - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 22-Jun-1999
C:Accession: 151663
R:Paterson, D.; Shi, Y.B.
J. Biol. Chem. 269, 25328-25334, 1994
A:Title: Thyroid hormone-dependent differential regulation of multiple arginase
genes during amphibian metamorphosis
A:Reference number: 151663; M01D:95014323
A:Accession: 151663
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-360 <PART>
A:Cross-references: EMBL:008406; NID:g497227; PIDN:AAA56893.1; PID:g497228
C:Superfamily: arginase 1

15163 Length: 360 February 11, 2000 15:51 Type: P Check: 4321 ..
1 MSIRSNFVRL LKKOVSIIKL QKCSHSVAV IGAPFSKQK RRGVEHGPAA
51 IRSAGLIERL SNLCNVCFD GDHFSQVFN DELYSIVKH PRYGLACKY
101 LAEVSRAVG AGHTCVTLGG DSHLAFSGIT GHAQCCPDIC VIWDAHADI
151 NTPLTPSGN LHGQVPSFL RELQDKVPI PGFSMAKPC LSKSDIYVIGL
201 RDLDPAEQFI LKNYISYS MRHIDCMGIR KWEKTFDOL LGRDRPIHL
251 SFDDIAPDPA LAPATGPVI GGLTYREGV ITEEIHNTGM LSAVDLEVN
301 PVLAATSEV KATANLAVDV IASCFQTR EGAHTRADTII DVLPTPSTY
351 ESDNEQVRI

11AA_SEQUENCE 1.0
P1:151664 - arginase 2 - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 22-Jun-1999
C:Accession: 151664
R:Paterson, D.; Shi, Y.B.
J. Biol. Chem. 269, 25328-25334, 1994
A:Title: Thyroid hormone-dependent differential regulation of multiple arginase
genes during amphibian metamorphosis
A:Reference number: 151664; M01D:95014323
A:Accession: 151664
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-360 <PART>
A:Cross-references: EMBL:008407; NID:g497229; PIDN:AAA56892.1; PID:g497230
C:Superfamily: arginase 1

15164 Length: 360 February 11, 2000 15:51 Type: P Check: 4641 ..
1 MSIRSNFVRL LKKOVSIIKL QKCSHSVAV IGAPFSKQK RRGVEHGPAA
51 IRSAGLIERL SNLCNVCFD GDHFSQVFN DELYSIVKH PRYGLACKY
101 LAEVSRAVG AGHTCVTLGG DSHLAFSGIT GHAQCCPDIC VIWDAHADI
151 NTPLTPSGN LHGQVPSFL RELQDKVPI PGFSMAKPC LSKSDIYVIGL
201 RDLDPAEQFI LKNYISYS MRHIDCMGIR KWEKTFDOL LGRDRPIHL
251 SFDDIAPDPA LAPATGPVI GGLTYREGV ITEEIHNTGM LSAVDLEVN
301 PVLAATSEV KATANLAVDV IASCFQTR EGAHTRADTII DVLPTPSTY
351 ESDNEQVRI

11AA_SEQUENCE 1.0
P1:151665 - arginase 3 - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 22-Jun-1999
C:Accession: 151665
R:Paterson, D.; Shi, Y.B.
J. Biol. Chem. 269, 25328-25334, 1994
A:Title: Thyroid hormone-dependent differential regulation of multiple arginase
genes during amphibian metamorphosis
A:Reference number: 151665; M01D:95014323
A:Accession: 151665
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-360 <PART>
A:Cross-references: EMBL:008408; NID:g497231; PIDN:AAA56893.1; PID:g497232
C:Superfamily: arginase 1

15165 Length: 360 February 11, 2000 15:51 Type: P Check: 4724 ..
1 MSIRSNFVRL LKKOVSIIKL QKCSHSVAV IGAPFSKQK RRGVEHGPAA
51 IRSAGLIERL SNLCNVCFD GDHFSQVFN DELYSIVKH PRYGLACKY
101 LAEVSRAVG AGHTCVTLGG DSHLAFSGIT GHAQCCPDIC VIWDAHADI
151 NTPLTPSGN LHGQVPSFL RELQDKVPI PGFSMAKPC LSKSDIYVIGL
201 RDLDPAEQFI LKNYISYS MRHIDCMGIR KWEKTFDOL LGRDRPIHL
251 SFDDIAPDPA LAPATGPVI GGLTYREGV ITEEIHNTGM LSAVDLEVN
301 PVLAATSEV KATANLAVDV IASCFQTR EGAHTRADTII DVLPTPSTY
351 ESDNEQVRI

11AA_SEQUENCE 1.0
P1:J50609 - biastictidin-S deaminase (EC 3.5.4.23) - Bacillus cereus
C:Species: Bacillus cereus
C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 26-Aug-1999
C:Accession: J50609; PS0264
R:Kobayashi, K.; Kamakura, T.; Tanaka, T.; Yamaguchi, I.; Endo, T.
Agric. Biol. Chem. 55, 3155-3157, 1991
A:Title: Nucleotide sequence of the bsr gene and N-terminal amino acid sequence
of biastictidin S deaminase from biastictidin S resistant Escherichia coli TK121.
A:Reference number: J50609; M01D:9214117
A:Accession: J50609
A:Molecule type: DNA
A:Residues: 1-140 <KOB>
A:Experimental source: strain K55-S1
A:Accession: PS0264
A:Molecule type: protein
A:Residues: 1-46 <KOB>
C:Genetics:
A:Gene: bsr
C:Superfamily: cytidine deaminase
C:Keywords: hydrolase

J50609 Length: 140 February 11, 2000 15:51 Type: P Check: 8425 ..
1 MKTFNISQDD LELVEVATEK IIMLYEDNKH HVGAAIRTKT GEISAVHIE
51 AYGRTVCA EALIGSAVS NGQKDPDTIV AVRRPYSDEV DSIRIVSPC
101 GKRCRELISDY APDCVLIEM NGKLVKTIE ELIPLKYTRN

11AA_SEQUENCE 1.0
P1:S37304 - spal protein - Salmonella typhimurium (fragment)
C:Species: Salmonella typhimurium
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 22-Jun-1999
C:Accession: S37304
R:Goisman, E.A.; Ochman, H.
EMBO J. 12, 3779-3787, 1993
A:Title: Cognate gene clusters govern invasion of host epithelial cells by
Salmonella typhimurium and Shigella flexneri.

```

A:Reference number: S37304; MUID:94008985  
A:Accession: S37304  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-335 <GRO>  
A:Cross-references: EMBL:X73525; NID:g404286; PIDN:CAAS1921.1; PID:g404287  
C:Genetics:  
A:Gene: spal  
C:Superfamily: H<sup>+</sup>-transporting ATP synthase alpha chain; H<sup>+</sup>-transporting ATP synthase alpha chain homology  
C:Keywords: P-loop  
F:63-70/Region: nucleotide-binding motif A (P-loop)  
F:86-254/Domain: H<sup>+</sup>-transporting ATP synthase alpha chain homology <ATP>  
S37304 Length: 335 February 11, 2000 15:51 Type: P Check: 2135 ..  
1 DPGKIYERF TPEVAPISE RVIDVAPPY ASRVGAREPL ITGVRAIDGL  
51 LTGCGGQRMG IFASAGCGKT MLMHMLEQT EADVFYIGLI GERGEVTEF  
101 VDMIRASHKK EKCULVFATS DFPYVDRCNA AQLATTVAEX FRDQCKRYVL  
151 FIDSMTRYAR ALRDVALASG ERPARRGYPA SVFDNLPRLI ERPGATSEGS  
201 ITAFYLVLE SEEDADPMAD EIRSLDQHL YLSRKLAGOG HYPALDIVKS  
251 VSRVFGQVTT PTHAEQASAV RKLMTREEL QLEFIDGEXR PGENDINDRA  
301 MOKRSLKAW LCPVAQYSS FDDTLGKNA FADQN  
11AA\_SEQUENCE 1.0  
P1:S01397 - H<sup>+</sup>-transporting ATP synthase (EC 3.6.1.34) chain I - thermophilic bacterium PS-3  
C:Species: thermophilic bacterium PS-3  
C>Date: 07-Sep-1990 #sequence\_revision 07-sep-1990 #text\_change 22-Jun-1999  
C:Accession: S01397  
R:Ohta, S.; Yohda, M.; Ishizuka, M.; Hirata, H.; Hamamoto, T.; Ohtawa-Hamamoto, Y.; Matsuda, K.; Kagawa, Y.  
Biochim. Biophys. Acta 933: 141-155, 1988  
A:Title: Sequence and over-expression of subunits of adenosine triphosphate synthase in thermophilic bacterium PS3  
A:Reference number: S01397; MUID:88163679  
A:Accession: S01397  
A:Molecule type: DNA  
A:Residues: 1-127 <OHT>  
A:Cross-references: EMBL:X07804; NID:g45808; PIDN:CAA30647.1; PID:g45809  
C:Superfamily: Bacillus H<sup>+</sup>-transporting ATP synthase chain I  
C:Keywords: ATP biosynthesis; hydrolase  
S01397 Length: 127 February 11, 2000 15:51 Type: P Check: 4392 ..  
1 MGNLQAMFWR QVRILYLLA IYTGFGFTP YKTVFSLIL GTSISLKW  
51 NLMTIKKFG QAVAAKKVR TLGLSLRAL AALAVIVLT YQGFHIVPT  
101 VLGLMYSIV IILDFEFHKW KNDLQA  
11AA\_SEQUENCE 1.0  
P1:S22348 - H<sup>+</sup>-transporting ATP synthase (EC 3.6.1.34) delta chain precursor - human  
C:Species: Homo sapiens (man)  
C>Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 22-Jun-1999  
C:Accession: S22348; T00493  
R:Jordan, E.M.; Breen, G.A.M.  
Biochim. Biophys. Acta 1130, 123-126, 1992  
A:Title: Molecular cloning of an import precursor of the delta-subunit of the human mitochondrial ATP synthase complex.  
A:Reference number: S22348; MUID:92182007  
A:Accession: S22348  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-168 <OOR>  
A:Cross-references: EMBL:X63422; NID:g12585; PIDN:CAA45016.1; PID:g12586  
R:Jamerdin, J.E.; McCreedy, P.M.; Skowronski, E.; Adamson, A.M.; Birkhart-Schultz, K.; Gordon, L.; Kyle, A.; Ramirez, M.; Stilaengen, S.; Phan, H.; Velasco, N.; Gaines, J.; Danganan, L.; Poundstone, F.; Christensen, M.; Georgescu, A.; Avila, J.; Liu, S.; Attix, C.; Andreise, T.; Tranknein, M.; Amico-Keller, G.; Goelfeld, J.; Duarte, S.; Lucas, S.; Bruce, R.; Thomas, P.; Qian, G.; Krommiller, B.; Ariellano, A.; Montgomery, M.; Ow, D.; Nolan, M.; Trong, S.; Kobayashi, A.; Olsen, A.O.; Carrano, A.V.  
submitted to the EMBL Data Library, February 1998  
A:Description: Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a serine protease gene cluster.  
A:Reference number: 214157  
A:Accession: T00493  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-168 <LAM>  
A:Cross-references: EMBL:AC004221; NID:g2911257; PIDN:AC004304.1; PID:g2911258  
C:Genetics:  
A:Gene: ATP5D  
A:Map position: 19  
A:Introns: 47/3; 99/1; 128/3  
C:Complex: the ATP synthase F1 complex consists of three alpha chains (see PIR:FMH0D), three beta chains (see PIR:A33370), one gamma chain (see PIR:A49108), one delta chain, and one epsilon chain; the F1 complex binds to the intrinsic membrane F0 complex  
C:Function:  
A:Description: catalyzes the formation of ATP from ADP and phosphate using the free energy derived from proton transport down the gradient maintained by cytochrome-c oxidase across the mitochondrial inner-membrane  
A:Pathway: oxidative phosphorylation  
C:Superfamily: H<sup>+</sup>-transporting ATP synthase epsilon chain  
C:Keywords: hydrolase; mitochondrion  
S22348 Length: 168 February 11, 2000 15:51 Type: P Check: 1686 ..  
1 MPEALLRRP GLGRVRRR AYAEAAAPA AASGNGNSF TFASPROVFF  
51 NGANVQGVVY PLUGARGIL AAHVPLQYL RGLGVVVAE DGTSKYFVS  
101 SSGIANNADS SVQLAEAV TUDMLDGA KANLEKQAE LVGTADENR  
151 AEQIRIEAN EALYKALE  
11AA\_SEQUENCE 1.0  
P1:D7121 - ATP synthase chain D - Chlamydia pneumoniae (strain CWL029)  
C:Species: Chlamydia pneumoniae  
C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 26-Aug-1999  
C:Accession: D7121  
R:Kelman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.; Davis, R.W.; Stephens, R.S.  
Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000; MUID:99206606  
A:Accession: D7121  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-209 <ARN>  
A:Cross-references: GB:AE001594; GB:AE001363; NID:g4376341; PID:g4376349  
A:Experimental source: strain CWL029  
C:Genetics:  
A:Gene: atpD  
C:Superfamily: H<sup>+</sup>-transporting ATPase chain D  
D7121 Length: 209 February 11, 2000 15:51 Type: P Check: 2102 ..  
1 MSYQVKLRN SFLEKOLA RLOTYVPTLK LKALLQAEV QNANVDAEC  
51 DKDYQVAYR IYARAEFSI PLCTDVEKS FEIOSINDF ENTAGVEVPI  
101 VREVTLFPAS YSLGTPIML DTMLSASKEL VVKKMAEVS KERLKLILEE  
151 LRAVSIRVNL FEKKLIPETI KIKKIAVFL SDRSITDVGQ VKNAKKIEL  
201 RKARGDECV

```

11AA_SEQUENCE 1.0
P1:S31498 - ribulose-bisphosphate carboxylase (EC 4.1.1.39) small chain - potato
C:Species: Solanum tuberosum (potato)
C>Date: 03-Mar-1994 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
C:Accession: S31498
R:Fitiz, C.C.; Wolter, F.P.; Schenkemeyer, V.; Herget, T.; Schreier, P.H.
Submitted to the EMBL Data Library, December 1992
A:Description: The gene-family of ribulose-(1,5) bisphosphate
carboxylase/oxygenase small subunit from Potato.
A:Reference number: S31494
A:Accession: S31498
A:Molecule type: DNA
A:Residues: 1-181 <P1>
A:Cross-references: EMBL:X69763; NID:g21570; PIDN:CAA9417.1; PID:g21571
C:Genetics: 60/3; 105/3
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
C:Superfamily: ribulose-bisphosphate carboxylase small chain
C:Keywords: carbon-carbon lyase; carboxy-lyase; chloroplast
S31498 Length: 181 February 11, 2000 15:51 Type: P Check: 4433

1 MASSIVSAA VATRSVNAQA SMVAFPTGLK SAASFVYTK NNNVDITSLA
51 SNGGRVRCMQ VMPPIMKRY ETLSTLPDL DEQLKEVEY LKNGWVPC
101 ETEHEGFVY RENHKSFGY DGRWTMVKL PMFGCTDAPQ VLAEOEAKK
151 AYPQAWIRII GFDNRVOYC ISFIAYKPEG Y

11AA_SEQUENCE 1.0
P1:A71295 - probable alanine racemase (alr) - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 26-Aug-1999
C:Accession: A71295
R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson,
R.; Gwin, M.; Hickey, E.R.; Clayton, R.; Ketchum, K.A.; Sodergren, E.;
Harcham, J.M.; McLeod, M.P.; Salzman, S.; Peterson, J.; Khalak, H.;
Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDonald, L.;
Artach, P.; Bowman, C.; Cotton, M.D.; Fujii, C.; Garland, S.; Hatch, B.;
Horst, K.; Roberts, K.; Matthey, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis
spirochete.
A:Reference number: A71250; NCID:98332770
A:Accession: A71295
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-357 <COL>
A:Cross-references: GB:AE001242; GB:AE000520; NID:g3322976; PID:g3322978
A:Experimental source: strain Nichols
C:Genetics: TP0681
A:Gene: TP0681
C:Superfamily: alanine racemase
A71295 Length: 357 February 11, 2000 15:51 Type: P Check: 255

1 MSRTARVCL PYKADYVGHG ACQVAQALS CGVSEFVAC VOEASOLRAA
51 GVRAPIILCS TPTAEISSL IEHRVHTVS ERAHIALIAR ALROSADTGA
101 TCGVHVKITD GMRIGCAPD EACALVOWC ATPGLHEGY CTHEFVADSV
151 RAEDLOYTEM QRAHFMCVQ YIRKSGISIP LVHANSAL LCHPRAHEDM
201 VRPGILAVGY APESVHPAVR SVELFVMELV TOVRAIKIP ACAYVSYQRL
251 WRAHETHVHG ILPIGVADGV MRALSPGLQV CIGKWPVAV GACIMDQCV
301 DIGTLARTV GDRVTLFPGQ DAGPGQAGAD VLAHSHATIP YELCALIGKR
351 VERVYIR

11AA_SEQUENCE 1.0
P1:G70553 - probable echA10 protein - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Jun-1999
C:Accession: G70553
R:Coile, S.T.; Brosch, R.; Parhill, J.; Garnier, T.; Churcher, C.; Harris, D.;
Gordon, S.V.; Eiglmeyer, K.; Gas, S.; Barry III, C.E.; Tekale, F.; Badcock, K.;
Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagers, K.;
Krogh, A.; McLean, J.; Mole, S.; Murrey, L.; Oliver, S.; Osborne, J.; Quail,
M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.;
Squires, S.
Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.
A:Reference number: A70500; NCID:98295987
A:Accession: G70553
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-268 <COL>
A:Cross-references: GB:Z9584; GB:AL123456; NID:g3261774; PIDN:CAB09030.1;
PID:e317233; PID:g2117180
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: echA10
C:Superfamily: naphthate synthase; enoyl-CoA hydratase homology
E_135-187/Domain: enoyl-CoA hydratase homology <ECH>
G70553 Length: 268 February 11, 2000 15:51 Type: P Check: 8877

1 MSNYRIDRT IVEGLAVTLA DGVLSVTDR PESINSITRK VLAGNADIE
51 GAATPRVKV VRUGAGRGF SGGALISVD VMASGPTDT VAEANRYTRA
101 IYALPQVYA VQGPYVCG VSLALADLV LASDNAFEML AHTVNGLMPD
151 GGASALVOAA IGRIRAHMA LIPDRVPAE ALSMGLVASV YPADFDIAEV
201 DKLISRLLAG PALAIAKTKN AINAATLET APTLLRELDG QALLRTDDE
251 AEGATAFOOR RTPEFTGR

11AA_SEQUENCE 1.0
P1:H69457 - ornithine cyclodeaminase (arcB) homolog - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 26-Aug-1999
C:Accession: H69457
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum,
K.A.; Dodson, R.J.; Gwin, M.; Hickey, E.R.; Peterson, J.D.; Richardson, D.L.;
Kerlavage, A.R.; Graham, D.E.; Kyrleides, N.C.; Fleischmann, R.D.; Quackenbush,
J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirsch, E.F.; Dougherty, B.A.;
McKenney, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.;
Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.;
McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Overbeek, R.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.;
Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;
Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic,
sulfate-reducing archaeon Archaeoglobus fulgidus.
A:Reference number: A69250; NCID:98049343
A:Accession: H69457
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-322 <KLE>
A:Cross-references: GB:AE000988; GB:AE000782; NID:g2689311; PID:g2648890;
TIGR:AF1665
C:Superfamily: ornithine cyclodeaminase
H69457 Length: 322 February 11, 2000 15:51 Type: P Check: 4102

1 METLITQEE VESLISMDFA MNAVEAFRL YALGKAOMP KYLEFEKGD

```

51 LRAMPALHMG YAGLKWNSH PGNPDGRLPT VVALMILNSP ETGEPILAVMD  
 101 ATYTSLRTG AAGGIAAKYL ARKNSVFGF ICGGTQAVQ LEALRRVFDI  
 151 GGVKADVRE KAKKRVSYC EDRGISASVQ PAEASRCDV LVTTSPSRKP  
 201 VYKAEWVEG THINAGADG PKRQELDVEI LKAKIYVDV LEQAKHGGEI  
 251 NVAWSGVIC VEDVHATIGE VINGLDGRE SDEITIFDS TGLAQGVAV  
 301 ARVYENALMS KNVSKIKF RI

# !!AA:SEQUENCE 1.0

PID:575089 - UDP-glucose 4-epimerase - *Synechocystis* sp. (strain PCC 6803)  
 N:Alternate names: protein S110244  
 C:Species: *Synechocystis* sp.

A:Variety: PCC 6803  
 C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 16-Jul-1999  
 C:Accession: S75089  
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hikosawa, M.; Sugiyura, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsumo, A.; Muraki, A.; Nakazaki, N.; Nanno, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.  
 DNA Res. 3, 109-136, 1996  
 A>Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.  
 A:Reference number: S74322; MUID:97061201  
 A:Accession: S75089  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-338 <K>A>  
 A:Cross-references: EMBL:D90910; GB:AB001339; NID:J1652956; PIDN:BAH1951.1; PID:010864; PID:91653034  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:  
 A:Gene: galE  
 A:Superfamily: *Escherichia coli* UDPglucose 4-epimerase; UDPglucose 4-epimerase homology  
 F:7-332/Domain: UDPglucose 4-epimerase homology <UDP>

S75089 Length: 338 February 11, 2000 15:51 Type: P Check: 8968 ..

1 MATOQTIIYT GGAGYISGSHG VIALQAGFD VLIYDNLSTG HKELVOPLGV  
 51 ELVVGHTGDR OKLDQLEFTR NIAAVMHFAA FIAGVESYOK PDIYQNNVV  
 101 GTTLLEAML AAGIKRFVFS STCAVYGMPK EIPMTESHPRI DPLSPYASK  
 151 RAVEQIADF DQAYGKFSVI FRFENSGAD PQRGLGEDHN PETHLPLAL  
 201 LRAKQRPQL SVFGTDYDTL DGLALRDYIH VCDLAIANVL GLQYLLGGE  
 251 SNIFNNGN GFSYQVLEV AKAVTGIDLP YQLCRRRGD APLVSSAK  
 301 ARELIGMFO YPDLHTIIDH AMQWQRRHG DYKSKNKI

!!AA:SEQUENCE 1.0  
 PID:H69105 - dtp-glucose 4,6-dehydratase - *Methanobacterium thermoautotrophicum* (strain Delta H)  
 C:Species: *Methanobacterium thermoautotrophicum*  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 16-Jul-1999  
 C:Accession: H69105  
 R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Basistrade, R.; Blakey, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lumu, W.; Potlifer, B.; Qiu, D.; Spadafora, R.; Viclaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Garuso, A.; Bush, D.; Sifer, H.; Patwell, D.; Prabhakar, S.; McDougall, S.; Shiner, G.; Goyal, A.; Pietrowski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
 J. Bacteriol. 179, 7155-7155, 1997

A>Title: Complete genome sequence of *Methanobacterium thermoautotrophicum* Delta H: functional analysis and comparative genomics.  
 A:Reference number: A69000; MUID:98037514  
 A:Accession: H69105

A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-336 <MTH>  
 A:Cross-references: GB:AE000933; GB:AE000666; NID:92622908; PIDN:AA86255.1; PID:92622920  
 A:Experimental source: strain Delta H  
 C:Genetics:  
 A:Gene: MTH1789  
 C:Superfamily: *Escherichia coli* UDPglucose 4-epimerase; UDPglucose 4-epimerase homology  
 F:4-314/Domain: UDPglucose 4-epimerase homology <UDP>

H69105 Length: 336 February 11, 2000 15:51 Type: P Check: 3451 ..

1 MEKILVTGGA GFISNFIRY MQEHPYHII NDLALTYCGN LBNLRGVEDE  
 51 PRYFVRGSI TDRRLVDGII KQVDAYINFA AESHVDSIE DPEIFRTNI  
 101 LCTQILLEAS RKRGVERFIQ ISTDEVYGA EKGYTEETP LABNSPYAS  
 151 KASADLVNRA YHRTYGLPVN ITRCSNNYGP YQPEKLIPL MITNALENRP  
 201 LPVYDGMNV RDMTHVLDC RAVDLVLRG RVGEVYNGG NSERNIEIV  
 251 ELVRELKGD ESTLRFVEDR PGHRRYARD ASKIRNELGW KPLYSFEEGI  
 301 RETIRMYIDN RDMWENIKSG EYLRYERYMY GGRLOD

!!AA:SEQUENCE 1.0  
 PID:D69290 - dtp-glucose 4,6-dehydratase (rfbB) homolog - *Archaeoglobus fulgidus*  
 C:Species: *Archaeoglobus fulgidus*  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 16-Jul-1999  
 C:Accession: D69290

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.V.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kierlavage, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirsch, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.T.; McNeil, L.K.; Badger, J.R.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
 Nature 390, 364-370, 1997  
 A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A>Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus*.  
 A:Reference number: A69250; MUID:98049343  
 A:Accession: D69290  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-332 <KLE>  
 A:Cross-references: GB:AB001082; GB:AE000782; NID:92689405; PIDN:AA890911.1; PID:92650312; TIGR:AF0324  
 C:Superfamily: *Escherichia coli* UDPglucose 4-epimerase; UDPglucose 4-epimerase homology  
 F:3-314/Domain: UDPglucose 4-epimerase homology <UDP>

D69290 Length: 332 February 11, 2000 15:51 Type: P Check: 128 ..

1 MRILVTGGG FIGSNEFVYA LAHQVEILN IDALKYGNP ENLKDLAEDE  
 51 RRSFKIGDIN DYELVSDIK KYDAVNFAA ESHVDRSIS PYAFIESNFL  
 101 GYITILEAVR KCNPEVRLV VSTDEVYGI EKSFTEDKG LMSPPYSAS  
 151 KAAQMLVIG YARTYNLAS ITRCTNNYGP YQPEKLIPL TIIRAMNLS  
 201 VPIYGTGMV RDMTHVLDC RAVEMVMEG ERREYNISS GEERTNLEIV

251 KTLLEIMGKD ESLIETVEDR PGHDVRSUD SWKIMRELKW RPKISFEEGI  
 301 RTIVETVLON EMMWRPLADE RILHPTPKL RV

!!AA\_SEQUENCE 1.0  
 P1:547045 - UDPglucose 4,6-dehydratase (EC 4.2.1.46) - Neisseria gonorrhoeae  
 C:Species: Neisseria gonorrhoeae  
 C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 16-Jul-1999  
 C:Accession: S47045  
 R:Robertson, B.D.; Frosch, M.; van Putten, J.P.  
 Submitted to the EMBL Data Library, April 1994  
 A:Description: Identification and characterisation of cryptic rhamnose  
 biosynthesis genes in Neisseria gonorrhoeae.  
 A:Reference number: S47045  
 A:Accession: S47045  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-346 <ROB>  
 A:Cross-references: EMBL:Z32742; NID:g520895; PIDN:CAB83652.1; PID:g520896  
 C:Gene: rfbB  
 C:Superfamily: Escherichia coli UDPglucose 4-epimerase; UDPglucose 4-epimerase  
 homology  
 C:Keywords: carbon-oxygen lyase; hydro-lyase  
 F:3-343/Domain: UDPglucose 4-epimerase homology <UDP>  
 S47045 Length: 346 February 11, 2000 15:51 Type: P Check: 5086 ..

1 MOTCKKNIL VTGAGFIS AVVHITQNT ROSVNLDKL TYAGNLESLT  
 51 DIADNPRIAF EOVDCIDRAE LDRVFAQYR DAVMHLAAS HYDAISAG  
 101 EEIFNTIVGT EDLEEARAY WQMPSEKRE AFRFHISTD EYGDHLGTD  
 151 DLFTETTPYA PSSPSASKA AADHLVRAMQ RTYRPSIYS NCSNNYGRO  
 201 FPEKLIPIMI LNALSGRPLV YYGDAQDIR MFEVEDHARA LYQVTEGIV  
 251 GETYNIGGN EKTNELEVNT ICALLEELAP EKPAVAYE DLITFYODRP  
 301 GHDAVAVDA AKIRDLGWL PLETFESGLR KIVQWYLDNK TRONA

!!AA\_SEQUENCE 1.0  
 P1:E70566 - Probable DRDP-GLUCOSE 4 - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 16-Jul-1999  
 C:Accession: E70566  
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Teale, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Hornby, T.; Jagers, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; 537-544, 1998  
 Nature 393, 537-544, 1998  
 A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence.  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: E70566  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-331 <COL>  
 A:Cross-references: GB:295390; GB:AL123456; NID:g3261766; PIDN:CAB08730.1;  
 PID:g316065; PID:g2104387  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: rmlB  
 C:Superfamily: Escherichia coli UDPglucose 4-epimerase; UDPglucose 4-epimerase  
 homology  
 F:3-315/Domain: UDPglucose 4-epimerase homology <UDP>  
 E70566 Length: 331 February 11, 2000 15:51 Type: P Check: 5640 ..

1 MRLIVTGAG FIGTNFVSHA VREHPDDAVT VDALTYAGR RESIADVEDA  
 51 IRLVGDITD AELVASOLVA SDAVHFAE SHVDNALNP EPLHNTIVIG  
 101 TFLLEAVR HGVRLHIST DEVYGDLELD DARETESTP YNPSPYSAT  
 151 KAGADMVRA WRSYGVAT ISNCSNNYGP YGHVEKFIPI QINVLIGRR  
 201 PKLYGAANY RMIVHDDN SAVRIIDRG RIGRHYLLSS EGEDNITVL  
 251 RTILRLMDR PDDEHVDYR VGHRLRAID PSTLDELQW APRHTDEEG  
 301 LRTITWYRD NESWRPLKD ATEARIERG Q

!!AA\_SEQUENCE 1.0  
 P1:G70415 - nucleotide sugar epimerase - Aquifex aeolicus  
 C:Species: Aquifex aeolicus  
 C>Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 16-Jul-1999  
 C:Accession: G70415  
 R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Overbeek, R.; Sneed, M.A.; Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short, J.M.; Olson, G.J.; Swanson, R.V.  
 Nature 392, 353-358, 1998  
 A:Title: The complete genome of the hyperthermophilic bacterium Aquifex  
 aeolicus.  
 A:Reference number: A70300; MUID:98196666  
 A:Accession: G70415  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-321 <NOF>  
 A:Cross-references: GB:AE000735; NID:g2983749; PIDN:AC07310.1; PID:g2983750;  
 GB:AE000657  
 A:Experimental source: strain VFS  
 C:Genetics:  
 A:Gene: nse  
 C:Superfamily: Escherichia coli UDPglucose 4-epimerase; UDPglucose 4-epimerase  
 homology  
 F:3-316/Domain: UDPglucose 4-epimerase homology <UDP>  
 G70415 Length: 321 February 11, 2000 15:51 Type: P Check: 2675 ..

1 MNILVTCAG LIGKVSXL LEHGRVIGV DNINSYDPR LKERYLEOLK  
 51 KRENKFKYV NIENREALRI LFOEFEDAV INEABAGVR YSIONPHIV  
 101 TTNLTGNLNL LELMKFEGVK KLILASTSSL YAGQPMFKE ELVNTPIISP  
 151 YVASKRAEV TAYTHYLXG IDVILRYFT VYGLAGRPDM AVFNFIYKTL  
 201 KQIPKRYGD GSQRDEFTY DVVAEATYKA LNKGYEIIIN VGNKPRALK  
 251 ELIELIKYT GKQVKEVYGD FHKADMDTW ADITKAKRLL GWEPKTSIEE  
 301 GYKKTVEWFL ENMDWVKDLR V

!!AA\_SEQUENCE 1.0  
 P1:E6750 - glucosamine-6-phosphate isomerase homolog ybf - Bacillus subtilis  
 C:Species: Bacillus subtilis  
 C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 16-Jul-1999  
 C:Accession: E6750  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Bertero, M.G.; Bessières, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Bouziers, L.; Brans, A.; Braun, M.; Brignelli, S.C.; Bron, S.; Broutillet, S.; Brunsch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denicola, F.; Devigne, K.M.; Duesterhoef, A.; Enllich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizti, A.; Galleron, N.; Gilm, S.Y.; Glaser, P.; Goffeau, A.; Gollightly, E.J.; Grandi, G.; Guisepi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.;

Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.

A:Authors: Lamber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Mostl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Onaga, B.; Park, S.H.; Parro, V.; Polk, T.M.; Portetelle, D.; Porciliak, S.; Prescott, A.M.; Priesack, E.; Pujic, F.; Purnelle, B.; Rapoport, G.; Reij, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Seto, T.; Scanlon, E.

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.-J.; Seror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Yamakoshi, A.; Tanaka, T.; Tepstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, E.; Vassarotti, A.; Viari, A.; Wandurt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A:Reference number: A69580; MUID:98044033

A:Accession: E69750

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-249 <RDN>

A:Cross-references: GB:599105; GB:AL009126; NID:g2632457; PIDN:CA812030.1; PID:el182188; PID:g2632522

A:Experimental source: strain 168

A:Genetics: YD1T

C:Superfamily: glucosamine-6-phosphate isomerase

E69750 Length: 249 February 11, 2000 15:51 Type: P Check: 3388 ..

1 MKILAEHY ELCKLSAII KEOIAKKA VLGIATGTP VGIYKOLISD

51 YQAGEIDFSK VTTFNDEYA GLSPSHPOSY NHEHEHLFQ HINMPDHIT

101 IPOGDNFQLE AACKYEDLI RQAGIDVOI LGIGANGHIG FNEPQSDPED

151 RTRYVKSLES TIQANARFEG GDPVLVPLRA ISMGIKTIME FSKHYILLAS

201 GEEKADAIQK MEGPVTTDV PASILOKKNH VTIADYAKAA OKLKASFS

!!AA\_SEQUENCE 1.0

P1:D69094 - phosphoribosylaminoimidazolesuccinocarboxamide synthase -  
Methanobacterium thermoautotrophicum (strain Delta H)

C:Species: Methanobacterium thermoautotrophicum

C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 16-Jul-1999

A:Accession: D69094

R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lum, W.; Pochler, B.; Qiu, D.; Spadatore, R.; Vicalire, R.; Wang, Y.; Wleczkowski, J.; Gibson, S.; Jivani, N.; Caruso, A.; Bush, D.; Sater, H.; Patwell, D.; Prabhakar, S.; McDonnell, S.; Shiner, G.; Goyal, A.; Pietrovski, S.; Church, G.M.; Daniels, C.D.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.

J. Bacteriol. 179, 7135-7155, 1997

A:Title: Complete genome sequence of *Methanobacterium thermoautotrophicum* Delta H: functional analysis and comparative genomics.

H: functional analysis and comparative genomics.

A:Reference number: A69000; MUID:98037514

A:Accession: D69094

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-248 <MTH>

A:Cross-references: GB:AE000804; GB:AE000666; NID:g2621196; PIDN:AAB84676.1; PID:g2621212

A:Experimental source: strain Delta H

A:Genetics: KMH170

C:Superfamily: phosphoribosylaminoimidazolesuccinocarboxamide synthase

D69094 Length: 248 February 11, 2000 15:51 Type: P Check: 5550 ..

Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.

A:Authors: Lamber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Mostl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Onaga, B.; Park, S.H.; Parro, V.; Polk, T.M.; Portetelle, D.; Porciliak, S.; Prescott, A.M.; Priesack, E.; Pujic, F.; Purnelle, B.; Rapoport, G.; Reij, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Seto, T.; Scanlon, E.

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.-J.; Seror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Yamakoshi, A.; Tanaka, T.; Tepstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, E.; Vassarotti, A.; Viari, A.; Wandurt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A:Reference number: A69580; MUID:98044033

A:Accession: E69750

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-249 <RDN>

A:Cross-references: GB:599105; GB:AL009126; NID:g2632457; PIDN:CA812030.1; PID:el182188; PID:g2632522

A:Experimental source: strain 168

A:Genetics: YD1T

C:Superfamily: glucosamine-6-phosphate isomerase

E69750 Length: 249 February 11, 2000 15:51 Type: P Check: 3388 ..

1 MKILAEHY ELCKLSAII KEOIAKKA VLGIATGTP VGIYKOLISD

51 YQAGEIDFSK VTTFNDEYA GLSPSHPOSY NHEHEHLFQ HINMPDHIT

101 IPOGDNFQLE AACKYEDLI RQAGIDVOI LGIGANGHIG FNEPQSDPED

151 RTRYVKSLES TIQANARFEG GDPVLVPLRA ISMGIKTIME FSKHYILLAS

201 GEEKADAIQK MEGPVTTDV PASILOKKNH VTIADYAKAA OKLKASFS

!!AA\_SEQUENCE 1.0

P1:JX0206 - chymotrypsin inhibitor (Kunitz) WCI-3 precursor - winged bean

C:Species: *Psophocarpus tetragonolobus* (winged bean)

C>Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 16-Jul-1999

A:Accession: JX0206; JN0110; JX0049; S42564; S42563

R:Habu, Y.; Psachoknagul, S.; Umemoto, K.; Sakata, Y.; Ohno, T.

J. Biochem. 111, 249-258, 1992

A:Title: Structure and regulated expression of Kunitz chymotrypsin inhibitor genes in winged bean (*Psophocarpus tetragonolobus* (L.) DC.).

A:Reference number: JX0206; MUID:92234991

A:Accession: JX0206

A:Molecule type: DNA

A:Residues: 1-207 <HAB>

R:Psachoknagul, S.; Matsui, T.; Shibata, H.; Hara, S.; Ikenaka, T.; Okada, Y.; Ohno, T.

Plant Mol. Biol. 12, 51-58, 1989

A:Title: Sequence and expression of the mRNA encoding the chymotrypsin inhibitor in winged bean (*Psophocarpus tetragonolobus* (L.) DC.).

A:Reference number: JN0110

A:Accession: JN0110

A:Molecule type: mRNA

A:Residues: 1-207 <PEY>

A:Experimental source: immature seed

R:Shibata, H.; Hara, S.; Ikenaka, T.

J. Biochem. 104, 537-543, 1988

A:Title: Amino acid sequence of winged bean (*Psophocarpus tetragonolobus* (L.) DC.) chymotrypsin inhibitor, WCI-3.

A:Reference number: JX0049; MUID:89197853

A:Accession: JX0049

A:Molecule type: protein

A:Residues: 25-207 <SHI>

R:Habu, Y.; Sakata, Y.; Fukasawa, K.; Ohno, T.

Plant Mol. Biol. 23, 1139-1150, 1993

A:Title: Ubiquitous nuclear proteins bind to 5' upstream region of major Kunitz chymotrypsin inhibitor gene in winged bean.

A:Reference number: S42563; MUID:94122379

A:Accession: S42564

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-207 <HA2>

A:Cross-references: EMBL:DJ13976; NID:g218008; PIDN:BA03086.1; PID:d1003592;

1 DDXNIFLSP LSVSTAFAMT KLGACNDTLK QLMHEVEFEDT

!!AA\_SEQUENCE 1.0

P1:A19940 - antithrombin III - baboon (fragment)

C:Species: *Papio sp.* (baboon)

C>Date: 05-Jun-1987 #sequence\_revision 05-Jun-1987 #text\_change 12-Apr-1996

A:Accession: A19940

R:Stackhouse, R.; Chandra, T.; Robson, K.J.H.; Woo, S.L.C.

J. Biol. Chem. 258, 703-706, 1983

A:Title: Purification of antithrombin III mRNA and cloning of its cDNA.

A:Reference number: A19940; MUID:83108864

A:Accession: A19940

A:Molecule type: protein

A:Residues: 1-40 <STA>

C:Function: Inhibits in blood plasma thrombin and activated coagulation factor X, either weakly alone or strongly in the presence of heparin

C:Superfamily: antithrombin III

A19940 Length: 40 February 11, 2000 15:51 Type: P Check: 1681 ..

1 MDVKIDGPLY SGKAKDVLDT DDEIVAVRF RDDITADGE KKDTEMKGY

51 YNSYIAKIF EYLEAGVPT QYLELRPGC ILARKLEMP IEVIRNIIAA

101 GSIVRRFEPT EGGEFVPLPI QMDYKSDENG DPMNDIIL AGIATRDEL

151 EIKRIRLTHI NSVLRDFLKS RGLIIPDFKL EFGEDSSGRI RGDVSPDT

201 CRLDMEGGE PLDKDFPRRG EGVGVAVRR VARIIDDED IERNVVEL

!!AA\_SEQUENCE 1.0

P1:A19940 - antithrombin III - baboon (fragment)

C:Species: *Papio sp.* (baboon)

C>Date: 05-Jun-1987 #sequence\_revision 05-Jun-1987 #text\_change 12-Apr-1996

A:Accession: A19940

R:Stackhouse, R.; Chandra, T.; Robson, K.J.H.; Woo, S.L.C.

J. Biol. Chem. 258, 703-706, 1983

A:Title: Purification of antithrombin III mRNA and cloning of its cDNA.

A:Reference number: A19940; MUID:83108864

A:Accession: A19940

A:Molecule type: protein

A:Residues: 1-40 <STA>

C:Function: Inhibits in blood plasma thrombin and activated coagulation factor X, either weakly alone or strongly in the presence of heparin

C:Superfamily: antithrombin III

A19940 Length: 40 February 11, 2000 15:51 Type: P Check: 1681 ..



PID: g218009; EMBL: D13975; NID: g218006; PID: d1003591; PID: g218007  
 A: Note: the nucleotide sequence was submitted to the EMBL Data Library, ./.  
 C: Superfamily: trypsin inhibitor (Kunitz)  
 C: Keywords: serine proteinase inhibitor  
 F: 1-24/Domain: signal sequence #status predicted <SIG>  
 F: 25-207/Product: chymotrypsin inhibitor (Kunitz) WCI-3 #status experimental  
 <MAT>  
 F: 89/Inhibitory site: Leu (chymotrypsin) #status predicted

JX0206 Length: 207 February 11, 2000 15:51 Type: P Check: 7697 ..

1 MKSTFFLALF LLSAISHLP SSTADDDLVD AEGNLVENG TYLLPHITWA  
 51 HGGGIETAKT GNEPCPLTVV RSPNEVSKGE PIRISSQFLS LPIRGSLVA  
 101 LGFANPPSCA ASPMTWTVDS POGPAVKLSQ OKLPENDILV FKEFKVSHN  
 151 IHVKKLYCQ HDEEDVKCQD YIGIHRDNG NRRLVTEEN PLEVLTKAK  
 201 SETASSH

!!AA\_SEQUENCE 1.0  
 P1: JX0650 - chymotrypsin inhibitor (Kunitz) WCI-2 precursor - winged bean  
 C: Species: Psophocarpus tetragonolobus (winged bean)  
 C: Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 07-Feb-1997  
 C: Accession: JX0650  
 R: Hanu, Y.; Poyachoknagul, S.; Umemoto, K.; Sakata, Y.; Ohno, T.  
 J. Biochem. 121, 111-124, 1992  
 A: Title: Structure and regulated expression of Kunitz chymotrypsin inhibitor  
 genes in winged bean (Psophocarpus tetragonolobus (L.) DC.).  
 A: Reference number: JX0206; NUID: 92244991  
 A: Accession: JX0650  
 A: Molecule type: DNA  
 A: Residues: 1-207 <HAB>  
 C: Superfamily: trypsin inhibitor (Kunitz)  
 C: Keywords: serine proteinase inhibitor  
 F: 1-24/Domain: signal sequence #status predicted <SIG>  
 F: 25-207/Product: chymotrypsin inhibitor (Kunitz) #status predicted <MAT>  
 F: 89-90/Region: inhibitory

JX0650 Length: 207 February 11, 2000 15:51 Type: P Check: 8167 ..

1 MKSTFFLALF LLSAISHLP SSTADDDLVD AEGNLVENG TYLLPHITWA  
 51 HGGGIETAKT GNEPCPLTVV RSPNEVSKGE PIRISSQFLS LPIRGSLVA  
 101 LGFANPPSCA ASPMTWTVDS POGPAVKLSQ OKLPENDILV FKEFKVSHN  
 151 IHVKKLYCQ RDEEDVKCQD YIGIHRDNG NRRLVTEEN PLEVLTKAK  
 201 SETASSH

!!AA\_SEQUENCE 1.0  
 P1: JX0246 - serine proteinase inhibitor A precursor - arrowhead  
 C: Species: Sagittaria sagittifolia (arrowhead)  
 C: Date: 10-Jun-1993 #sequence\_revision 10-Jun-1993 #text\_change 16-Jul-1999  
 C: Accession: JX0246; JX0780  
 R: Xu, W.F.; Tao, W.K.; Gong, Z.Z.; Chi, C.W.  
 J. Biochem. 119, 153-158, 1993  
 A: Title: cDNA and genomic structures of arrowhead proteinase inhibitors.  
 A: Reference number: JX0246; MUID: 93224493  
 A: Accession: JX0246  
 A: Molecule type: DNA  
 A: Residues: 1-181 <XU1>  
 A: Cross-references: DDBJ: D13819; NID: g218280; PID: BAA02972.1; PID: d1003478;  
 PID: g218281  
 A: Note: the authors translated the codon GTC for residue 166 as Ile  
 A: Accession: JX0780  
 A: Molecule type: mRNA  
 A: Residues: 1-60, 'W', '62-83', 'S', '85-151', 'D', '153-173', 'F', '175-181 <XU2>  
 A: Note: 39-Arg, 142-Val, 143-Gln and 145-His were also found  
 C: Comment: This protein inhibits trypsin, chymotrypsin, elastase, and  
 lysylendopeptidase.

C: Genetics:  
 A: Introns: 121/2  
 C: Superfamily: arrowhead proteinase inhibitor B  
 C: Keywords: serine proteinase inhibitor  
 F: 1-24/Domain: signal sequence #status predicted <SIG>  
 F: 25-174/Product: serine proteinase inhibitor A #status predicted <MAT>

JX0246 Length: 181 February 11, 2000 15:51 Type: P Check: 1378 ..

1 MAASNALLI SGVLLISLAV LCHGDPVYDS DDAVQNLNG GNYPLTYIOS  
 51 AAIKRGGLS TLHKDCKSY VYEAPETDRG LPVGFASAT SQPMWQGSR  
 101 YKFSFMPVP LICDTAMSIG KSTETGVYK LAACSCFCK IACEVGSFN  
 151 VNGRTLGIQ GEHFTVRFQK FDALAKTAP Q

!!AA\_SEQUENCE 1.0  
 P1: JX5447 - serine proteinase inhibitor A precursor - arrowhead  
 C: Species: Sagittaria sagittifolia (arrowhead)  
 C: Date: 10-Jul-1997 #sequence\_revision 29-Aug-1997 #text\_change 07-May-1999  
 C: Accession: JX5447  
 R: Luo, M.J.; Lu, W.Y.; Chi, C.W.  
 J. Biochem. 121, 991-995, 1997  
 A: Title: Clarification of an uncertain intron within the cDNA sequences of  
 arrowhead proteinase inhibitors A and B.  
 A: Reference number: JX5447; NUID: 97335950  
 A: Accession: JX5447  
 A: Molecule type: mRNA  
 A: Residues: 1-210 <LUO>  
 C: Superfamily: arrowhead proteinase inhibitor B  
 C: Keywords: serine proteinase inhibitor  
 F: 1-24/Domain: signal sequence #status predicted <SIG>  
 F: 25-210/Product: serine proteinase inhibitor A #status predicted <MAT>

JX5447 Length: 210 February 11, 2000 15:51 Type: P Check: 4632 ..

1 MAASNALLI SGVLLISLAV LCHGDPVYDS DDAVQNLNG GNYPLTYIOS  
 51 AAIKRGGLS TLHKDCKSY VYEAPETDRG LPVGFASAT SQPMWQGSR  
 101 YKFSFMPVP LICDTAMSIG KSETNGISF QPITAGDYF LNFSEFAR  
 151 STEETGVYKL AACSCFCKI ACEVGSFNV NGRLLGIQ GEHFTVRFQK  
 201 DALAKTAPQ

!!AA\_SEQUENCE 1.0  
 P1: T06517 - alpha-amylase inhibitor 1ma1 precursor, monomeric - wheat  
 C: Species: Triticum aestivum (common wheat)  
 C: Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 16-Jul-1999  
 C: Accession: T06517  
 R: Carbonero, P.  
 submitted to the EMBL Data Library, February 1998  
 A: Reference number: 215731  
 A: Accession: T06517  
 A: Status: translated from GB/EMBL/DDBJ  
 A: Molecule type: mRNA  
 A: Residues: 1-151 <CAR>  
 A: Cross-references: EMBL: AJ223492; NID: e1251337; PID: CAA11410.1; PID: e1251338  
 C: Experimental source: cv. Chinese spring, endosperm  
 C: Genetics:  
 A: Gene: 1ma1  
 C: Superfamily: wheat alpha-amylase inhibitor  
 C: Keywords: alpha-amylase inhibitor  
 F: 1-30/Domain: signal sequence #status predicted <SIG>  
 F: 31-151/Product: alpha-amylase inhibitor 1ma1 #status predicted <MAT>

T06517 Length: 151 February 11, 2000 15:51 Type: P Check: 7537 ..

1 MMKTYFQGL LVPMVATFM AVEYGARSHN SGPMSCDDPA TGYKYSALNG  
 51 CRAMVKLCV GSQVPAVLK DCCQQLADIN NEMRCGDLS SMLRSYQEL

```
101 GVREGKEVLP GCRKEVYKLT AASVEVCV PIPNSGDRA GVCYMAAYD
151 V
!!AA_SEQUENCE 1.0
P1:S16920 - alpha-amylase inhibitor - wheat
C:Species: Triticum aestivum (common wheat)
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 31-Oct-1997
C:Accession: S16920
R:Caporale, C.; Carraro, L.; Nitti, G.; Poerio, E.; Pucci, P.; Buonocore, V.
Protein Seq. Data Anal. 4, 3-8, 1991
A:Title: Determination of the primary structure of an alpha-amylase inhibitor
from wheat kernel by Edman degradation and fast atom bombardment mass
spectrometry.
A:Reference number: S16920; MUID:92020810
A:Accession: S16920
A:Status: Preliminary
A:Molecule type: Protein
A:Residues: 1-121 <CAP>
C:Superfamily: wheat alpha-amylase inhibitor
S16920 Length: 121 February 11, 2000 15:51 Type: P Check: 2886 ..

1 SGPSWCDPA TGYKVSALTG CRAMVKLOV GSOVPEAVLR DCCOOLADIN
51 NEMCRGCDLS SMLRSYQEL GVREGKEVLP GCRKEVYKLT AASVEVCV
101 PIPNSGDRA GVCYMAAYD V
!!AA_SEQUENCE 1.0
F1:A60195 - transforming protein db1 - mouse (fragment)
N:Contains: transforming protein Mcf2
C:Species: Mus musculus (house mouse)
C:Date: 03-Mar-1993 #sequence_revision 17-Jul-1994 #text_change 07-May-1999
C:Accession: A60195; A61523
R:Galland, F.; Patis, Y.; delapeyriere, O.; Birnbaum, D.
Oncogene 6, 833-839, 1991
A:Title: Restriction and complexity of Mcf2 proto-oncogene expression.
A:Reference number: A60195; MUID:91270902
A:Accession: A60195
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-278 <GAL>
A:Note: authors translated the codon GAT for residue 229 as Ser
R:Grant, S.G.; Mattei, M.G.; Galland, F.; Stephenson, D.A.; Keltz, B.T.;
Birnbaum, D.; Chapman, V.M.
Cytogenet. Cell Genet. 54, 175-181, 1990
A:Title: Localization of the mouse Mcf-2 (db1) protooncogene within a conserved
linkage group on the mouse X chromosome.
A:Reference number: A61523; MUID:91092130
A:Accession: A61523
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-6, 'G', 8-73 <GR>
C:Superfamily: db1 transforming protein; CDC24 homology; plectstrin repeat
homology
C:Keywords: oncogene; transforming protein
F:113-161/Domain: CDC24 homology #status atypical <CD24>
F:119-278/Domain: plectstrin repeat homology (fragment) <PLK>
A60195 Length: 278 February 11, 2000 15:51 Type: P Check: 2543 ..

1 DQSPKLDNSL DILKNHYLNE LIOTERAYVR ELFTVLLGYR SEMDNPOMD
51 LMPPLLRKK DVLFGNMAEI YEFHNIFMS RLEDCSDAPE RVGPCFLERK
101 DDQFMVAKYC QNKPRSELIM RYSCAFQ ECQKRLKHL GDSYLLKRY
151 QRIRKYOLL KGNINELGKA VLQGSFNWL GHRKGATKK DARRPKMR
201 HLFYERAVM FCKRRESGE GADRPSPYDF KHCLMEDVG ITBHVAGNKR
251 KFEIWESEKE EIYIVAPNV DVKMLMK

!!AA_SEQUENCE 1.0
P1:A40316 - cyclin B - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 17-Jan-1992 #sequence_revision 17-Jan-1992 #text_change 16-Jul-1999
C:Accession: A40316
R:Buono, A.; Richardson, H.; Reed, S.I.; Russell, P.
Cell 66, 149-159, 1991
A:Title: A fission yeast B-type cyclin functioning early in the cell cycle.
A:Reference number: A40316; MUID:91300545
A:Accession: A40316
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-415 <RUE>
A:Cross-references: GB:M68881; NID:9173350; PIDN:AAA35288.1; PID:9173351
C:Superfamily: cyclin
C:Keywords: cell cycle control
A40316 Length: 415 February 11, 2000 15:51 Type: P Check: 7358 ..

1 MDVSTQTRRA TYFQDENOLQ KDHIYKKS HIKLNTGVA PERAVDINO
51 QDEPTLIEGN NESSISSSTG DFEEDFAYO DVEIEERSI RSTPSIGDD
101 DLENREGSFD APEGILTHGK HRLPTPEWT KEDLALSEA AARLOANPSP
151 EDIETDPSMV PDYDEIFHY MOSLERKLP PNYMSVOE IDWVTRMLV
201 DWIVQVQIHF RLLPETFLA VNLIDRFLSI KVSLOKVOL VGLSALLIAC
251 KYEIHPPSI YNEFAVVOGI FTVDEIRAE RYMLMLDRE ISWGPSPFL
301 RRSIRASVD HIRMLAKYL QEVTLMEIF ISAHSFIA TAYLSQML
351 GHDMTPCHV YSGYATROL KPCANIWEC LVDAPRHNA IYKYSNRM
401 KRVSAPFHHM VLSVI
!!AA_SEQUENCE 1.0
P1:A57234 - 11n-44 protein precursor - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 20-Sep-1999
C:Accession: A57234
R:Herman, M.A.; Vassiliava, L.L.; Horvitz, H.R.; Shaw, J.E.; Herman, R.K.
Cell 83, 101-110, 1995
A:Title: The Caenorhabditis elegans gene 11n-44, which controls the polarity of
certain asymmetric cell divisions, encodes a Wnt protein and acts cell
nonautonomously.
A:Reference number: A57234; MUID:96006529
A:Accession: A57234
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-348 <HER>
A:Cross-references: GB:U22179; NID:9758776; PID:9758777
C:Superfamily: Int-1 transforming protein
A57234 Length: 348 February 11, 2000 15:51 Type: P Check: 2377 ..

1 MRAAPDEFF OSTALSTFPI LCSLATNEIP TISGAPAGKI VQPKPNILK
51 QGCPSDLHS RALRSIOLAC RTHPATVISA FEGVOEGION CANLRFOOM
101 DCEAGNIMH DPPLRQGR ESSLIWASS ASAANGVATA CAQGMIDCA
151 CNNOGONEV EFGCTHGV HGTASRKL TWGAVENTL RYKRNILKA
201 GLAIKRTLI SSCCKHVSQ SCQKCTCKR TATLEHTDY LVEKYARAKL
251 YTDSDVYKTI DLIYLEASPD VCKAKSVAGR VCAMNEHTT QGDDRLCCG
301 NGSIRHEVY RYKCDCEFW CNLVCKDCI QHRWISTCNG TPKSLIF
```

P1:JC4152 - Wnt-11 protein precursor - chicken  
 C:Species: Gallus gallus (chicken)  
 C>Date: 27-Aug-1995 #sequence\_revision 19-Oct-1995 #text\_change 26-Aug-1999  
 C:Accession: J04152  
 R:Randa, N.; Ohuchi, H.; Yoshida, H.; Noji, S.; Nohno, T.  
 Biochem. Biophys. Res. Commun. 211, 123-129, 1995  
 A:Title: A chicken Wnt gene, Wnt-11, is involved in dermal development.  
 A:Reference number: J04152; MUID:95298011  
 A:Accession: J04152  
 A:Molecule type: mRNA  
 A:Residues: 1-354 <TAB>  
 A:Cross-references: DDBJ:D31901; NID:9505349; PID:01007271; PID:9505350  
 A:Experimental source: embryo  
 C:Comment: This protein is a cysteine-rich protein, and it plays an important role as a signaling factor in the cell-cell interaction during embryogenesis and as an inductive signal for dermal development.  
 C:Genetics:  
 A:Gene: Wnt-11  
 C:Superfamily: Int-1 transforming protein  
 C:Keywords: glycoprotein; mesoderm  
 F:1-25/Domain: signal sequence #status predicted <SIG>  
 F:26-354/Product: Wnt-11 protein #status predicted <MAT>  
 F:40,90/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 J04152 Length: 354 February 11, 2000 15:51 Type: P Check: 4630 ..

1 MKPSPQFLA AFLSLILQIG ICYGIKIAL SKTPSSIALN QTOCHKOLEG  
 51 LVYSQVQLCR SNLELMQITL QAAREVIKTC RRTFSDMRWN CSSTELAPNY  
 101 LLDERTRE SAFVYALSA AISHTIARAC TTGDLPGCCG GPFGEPGP  
 151 GYRGGGADN LNYGLMGSK FSDAPMKMKK SSGQANKMLH LHNSEVGROY  
 201 LKASLEKCK CHGVSGSCI KTCWKGLQEL RDTALDLNKK YLSATKYVHR  
 251 PMGTAKLIV KDIDIRPYKE TELLYLOSSP DCCMKREKYG SHGQODQCN  
 301 KTSNGSDSCD LMCGRGYNP YMDKVERCH CKYHCCYVT CKKCRIVER  
 351 YVCK

!!AA\_SEQUENCE 1.0  
 P1:S34378 - wnt-11 protein - mouse  
 C:Species: Mus musculus (house mouse)  
 C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 26-Aug-1999  
 C:Accession: S34378  
 R:Denise, C.L.; Delaney, S.J.; Christiansen, J.H.; Walwright, B.J.  
 Submitted to the EMBL Data Library, January 1993  
 A:Description: A novel member of the murine wnt gene family (wnt-11) is expressed during mouse embryogenesis.  
 A:Reference number: S34378  
 A:Accession: S34378  
 A:Molecule type: mRNA  
 A:Status: preliminary  
 A:Residues: 1-354 <DEN>  
 A:Cross-references: EMBL:X70800; NID:9312829; PID:9312830  
 C:Superfamily: Int-1 transforming protein  
 S34378 Length: 354 February 11, 2000 15:51 Type: P Check: 1689 ..

1 MRARQVCEA LFLALALHNG VCYGRKVAL SKTPPALALN QTOCHKOLEG  
 51 LVSAQVQLCR SNLELMQITL HAARAMAKC RRFADMRWN CSSTELAPNY  
 101 LLDERTRE SAFVYALSA TISHTIARAC TSGDLPCCSC GPVPEPFGP  
 151 GNRWGGADN LSYGLMGAK FSDAPMKYK TSGQANKMLR LHNSEVGQDA  
 201 LRASLETCK CHGVSGSCI RTCKWGLQEL QDYAADIKTR YLSATKYVHR  
 251 PMGTAKLIV KDIDIRPYKD SELVYLOSSP DCCMKREKYG SHGQODQCN

301 KTSNGSDSCD LMCGRGYNP YMDKVERCH CKYHCCYVT CKKCRIVER  
 351 YVCK

!!AA\_SEQUENCE 1.0  
 P1:JH0687 - bone morphogenetic protein 21 precursor - African clawed frog  
 C:Species: Xenopus laevis (African clawed frog)  
 C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 16-Jul-1999  
 C:Accession: JH0687; S16244  
 R:Nishimatsu, S.; Suzuki, A.; Shoda, A.; Murakami, K.; Ueno, N.  
 Biochem. Biophys. Res. Commun. 186, 1487-1495, 1992  
 A:Title: Genes for bone morphogenetic proteins are differentially transcribed in early amphibian embryos.  
 A:Reference number: JH0687; MUID:92378616  
 A:Accession: JH0687  
 A:Molecule type: mRNA  
 A:Residues: 1-398 <NIS>  
 A:Cross-references: GB:X63425; NID:964583; PIDN:CA445019.1; PID:964584  
 A:Experimental source: oocyte  
 C:Superfamily: Inhibin  
 C:Keywords: glycoprotein  
 F:285-398/Product: bone morphogenetic protein 21 #status predicted <MAT>  
 F:137,202,237,340/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 JH0687 Length: 398 February 11, 2000 15:51 Type: P Check: 1620 ..

1 MVAGIHPILL LFFYQILISG CTGLVPEBK RYSESTNRS PQSQOVLQD  
 51 FELRLNMG LKRRPTGKN VVIPYMLD YHLNSQLAD DQGSSEVDYH  
 101 MERAASRANT VRSFHEESM EEIPESEKKT IORFFNLSS IPDELYTSS  
 151 ELRIREQVQ EPEKTDGSKL HRINIYIVK PAAASRGV VRLDTRLIH  
 201 HNSKWESEF VTPAITRWIA HKOPNGFV EYTHLDNDTN VPKRHVRISR  
 251 SLTDKGHP RIRPLVTFE HDKGHALHK ROKRQARHKO RKRLKSCRR  
 301 HPLVDFSDV GMDWIVAPP GYHAFYCHG CPEPLADHLN STNHAIVQTL  
 351 VNSVNTNPK ACCVPELSA ISMLYLDENE KVALKNYQDM VVEGCGCR

!!AA\_SEQUENCE 1.0  
 P1:JH0687 - bone morphogenetic protein 21 precursor - African clawed frog  
 C:Species: Xenopus laevis (African clawed frog)  
 C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 16-Jul-1999  
 C:Accession: JH0687; S16244  
 R:Nishimatsu, S.; Suzuki, A.; Shoda, A.; Murakami, K.; Ueno, N.  
 Biochem. Biophys. Res. Commun. 186, 1487-1495, 1992  
 A:Title: Genes for bone morphogenetic proteins are differentially transcribed in early amphibian embryos.  
 A:Reference number: JH0687; MUID:92378616  
 A:Accession: JH0687  
 A:Molecule type: mRNA  
 A:Residues: 1-398 <NIS>  
 A:Cross-references: GB:X63424; NID:964585; PIDN:CA445019.1; PID:964586  
 A:Experimental source: oocyte  
 C:Superfamily: Inhibin  
 C:Keywords: glycoprotein  
 F:285-398/Product: bone morphogenetic protein 21 #status predicted <MAT>  
 F:137,202,237,340/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 JH0687 Length: 398 February 11, 2000 15:51 Type: P Check: 9563 ..

1 MVAGIHPILL LFFYQILISG CTGLVPEBK RYSESTNRS PQSQOVLQD

51 FELRLSMEG LKRRPPGKN VVIPYMDL YHLHLAOLAA DEGTAMDFQ  
 101 MERAASRANT VRSFHEESM EEIPESREKT IQREFNLSS IPNEELVISA  
 151 ELRIFREVOQ EPESDSSKL HRINTYDIK PAAASRGPV VRLDTRLVH  
 201 HNESKWESEFD VTPAIAFWIA HKOPNHFVY EVTHLDNDKN VPKKHVRISR  
 251 SLTPKQDWP QIRPLVTFPS HDGKGHALHK ROKRQARHKQ RKRLKSSCRH  
 301 HPLVDFSDV GWNMDIYAPR GYHAFYCHGE CFPPLADHLN STNHALVQRL  
 351 VNSVNTNIPK ACCVPELSA ISMLYLDENE KVLKNTQDM VVEGCGCR

!!AA\_SEQUENCE 1.0  
 P1:S21299 protein DAF8 - eastern green mamba  
 C:Species: Dendroaspis angusticeps (eastern green mamba)  
 C:Date: 07-Dec-1994 #sequence\_revision 12-May-1995 #text\_change 16-Jul-1999  
 C:Accession: S21299  
 R:Duncanell, F.  
 Submitted to the EMBL Data Library, June 1990  
 A:Reference number: S21298  
 A:Accession: S21299  
 A:Status: Preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-81 <DUC>  
 A:Cross-references: EMBL:X53409; NID:962755; PIDN:CAA37485.1; PID:962756  
 C:Superfamily: snake toxin

S21299 Length: 81 February 11, 2000 15:51 Type: P Check: 8703 ..

1 MKTLLTLVL VTIYCDLAS TMICYSHKRP QPSATICEE KTCYKKSVMK  
 51 LPAIAGRCG GCPSEKMLVA IHCCSDKCN E

!!AA\_SEQUENCE 1.0  
 P1:A37045 thermostable direct hemolysin homolog TRH precursor - Vibrio  
 C:Species: Vibrio parahaemolyticus  
 C:Date: 31-Jan-1992 #sequence\_revision 31-Jan-1992 #text\_change 30-Sep-1993  
 C:Accession: A37045  
 R:Nishibuchi, M.; Taniguchi, T.; Misawa, T.; Khaemane-Jam, V.; Honda, T.;  
 Miwatani, T.  
 Infect. Immun. 57, 2691-2697, 1989  
 A:Title: Cloning and nucleotide sequence of the gene (trh) encoding the  
 hemolysin related to the thermostable direct hemolysin of Vibrio  
 parahaemolyticus.  
 A:Reference number: A37045; MUID:89339712  
 A:Accession: A37045  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 1-189 <NIS>  
 C:Superfamily: thermostable direct hemolysin

A37045 Length: 189 February 11, 2000 15:51 Type: P Check: 6756 ..

1 MKLKIYFAPS LLLASITSVS KSFALDPSI PFPSPGSEL LFVYRNTTIK  
 51 TESPVNAIVN DYWTNRNIKR KPYKSVHGS IFTTSGSKWL SAIITVNING  
 101 NNYTMAALSG YKDGSLTVFT KSEKTSLNOD FYSVKSFVND SESISPSINY  
 151 LDETPPEYFVN VEAYESGNH MFWMCISNKL SFDECSMSON

!!AA\_SEQUENCE 1.0  
 P1:A48942 - thermostable direct hemolysin-related hemolysin TRH2 - Vibrio  
 C:Species: Vibrio parahaemolyticus  
 C:Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 16-Jul-1999  
 C:Accession: A48942  
 R:Kishishita, M.; Matsuka, N.; Kumagai, K.; Yamasaki, S.; Takeda, Y.;  
 Nishibuchi, M.

Appl. Environ. Microbiol. 58, 2449-2457, 1992  
 A:Title: Sequence variation in the thermostable direct hemolysin-related  
 hemolysin (trh) gene of Vibrio parahaemolyticus.  
 A:Reference number: A48942; MUID:92384559  
 A:Contents: AT4  
 A:Accession: A48942  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-189 <KIS>  
 A:Cross-references: GB:M88112; NID:915305; PIDN:AAA27578.1; PID:915306  
 A:Note: sequence extracted from NCBI backbone (NCBIN:112089, NCBIPI:112090)  
 C:Superfamily: thermostable direct hemolysin

A48942 Length: 189 February 11, 2000 15:51 Type: P Check: 5895 ..

1 MKLKIYFAPS LLLVSIFIS KSFALDPSI PFPSPGSEL LFVYRNTTIK  
 51 TESPVNAIVE DYWTNRNIKR KPYKDYVQS VFTTAGSKWL SAIMTVNINIG  
 101 NNYTMAALSG YKDGSLTVFT KSEKTSLNOD FYSVKSFVND SESISPSINY  
 151 LDETPPEYFVT VEAYESGNH MFWMCISNKL SFDECSMSON

!!AA\_SEQUENCE 1.0  
 P1:J00472 - T-cell receptor beta chain (BTB4) - bovine  
 C:Species: Bos primigenius taurus (cattle)  
 C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-May-1997  
 C:Accession: J00472  
 R:Tanaka, A.; Ishiguro, N.; Shinagawa, M.  
 Submitted to JIPID, May 1990  
 A:Description: Sequence analysis of bovine T-cell receptor beta chain genes.  
 A:Reference number: J00472  
 A:Accession: J00472  
 A:Molecule type: mRNA  
 A:Residues: 1-135 <TAN>  
 A:Experimental source: T cell  
 C:Genetics:  
 A:Gene: BTB4  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: receptor

J00472 Length: 135 February 11, 2000 15:51 Type: P Check: 2831 ..

1 MISCHRODSL GAGRQALGL ALSSISSPAA VSKSGASVTI ECRALDFQAT  
 51 TVFWYRQFPK RGLIMATSN VGSADTYEAG YNKDKIPISO PDLRFSSTLV  
 101 TRVDPADSL YFCGVHTASH PGKMHFGPGT RLIVL

!!AA\_SEQUENCE 1.0  
 P1:J00473 - T-cell receptor beta chain (BTB1) - bovine  
 C:Species: Bos primigenius taurus (cattle)  
 C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 30-May-1997  
 C:Accession: J00473  
 R:Tanaka, A.; Ishiguro, N.; Shinagawa, M.  
 submitted to JIPID, May 1990  
 A:Description: Sequence analysis of bovine T-cell receptor beta chain genes.  
 A:Reference number: J00473  
 A:Accession: J00473  
 A:Molecule type: mRNA  
 A:Residues: 1-136 <TAN>  
 A:Experimental source: T cell  
 C:Genetics:  
 A:Gene: BTB1  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: T-cell receptor

J00473 Length: 136 February 11, 2000 15:51 Type: P Check: 3911 ..

1 MISHRODSL GATRKALGLA LSSISSPAA SKSGASVTI CRALDFQASS  
 51 MFWYRQFPKR GLVLMATSNB GTADTYEGY NKKKFIQSP DRTFSLMWI

```

101 SVDPDSSLY FCSARDSVNA DTOPYFGAG SKLTIVL
!!AA_SEQUENCE 1.0
P1:B53250 - class I histocompatibility antigen BOLA-AW10 alpha chain precursor
- bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 23-Jul-1999
C:Accession: B53250
R:Bersaud, A.; Kaushal, A.; Baldwin, C.L.; Clevers, H.; Young, J.R.; Kemp,
S.J.; Machuga, N.D.; Toye, P.G.; Teale, A.J.
Immunogenetics 33: 247-254, 1991
A:Title: Identification of expressed bovine class I MHC genes at two loci and
demonstration of physical linkage
A:Reference number: A53250; MUID:91224708
A:Accession: B53250
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-353 <BEN>
A:Cross-references: GB:M69206; NID:q163345; PIDN:AAA30638.1; PID:q289420
C:Superfamily: class I histocompatibility antigen; Immunoglobulin homology
F:210-275/Domain: Immunoglobulin homology <IMM>
B53250 Length: 353 February 11, 2000 15:51 Type: P Check: 6952

1 LLLSGVLVT ETRAGSHSMR YFTAVSRPG LGEPYLEVG YVDTQYVRF
51 DSDAENPRME PRARWEOEG PEYMDRETQR AKGNAOFFRV SLNLRGYN
101 OSEAGSHTLQ WMSGCYVGPD GRPRGFMOF GYDGRDYLAL NEDLRMTAV
151 ETMAQISRRK MEAGAEAVQ RNYLGRVCE WLRRYLENGR DTLRADPPK
201 AHVTRPIRG REVTLRCMAL GFPEIEISLT WORGEDQOT DMELVETRPS
251 GDNQFQWAA LTVSGEEQK YTCQVHEGL QEPPLTKWP POPSFLTWGI
301 IYGLVLVVT GAVAVGVIC MKKRSGERK TYICASSDS AGSDVSLTV
351 PKV
!!AA_SEQUENCE 1.0
P1:15609 - T-cell surface glycoprotein CD8 beta chain - chicken
C:Species: Gallus gallus (chicken)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 23-Jul-1999
C:Accession: 150609; S37237
R:Teegaskes, C.A.; Kong, F.K.; Paramithiotis, E.; Chen, C.L.; Ratcliffe, M.J.;
Davison, T.F.; Young, J.R.
J. Immunol. 154, 4485-4494, 1995
A:Title: Identification and analysis of the expression of CD8 alpha beta and
CD8 alpha alpha isoforms in chickens reveals a major TCR-gamma delta CD8 alpha
beta subset of intestinal intraepithelial lymphocytes.
A:Reference number: 150609; MUID:95238946
A:Accession: 150609
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-207 <TR>
A:Cross-references: EMBL:Z26484; NID:9403300; PIDN:CA81258.1; PID:9403301
C:Superfamily: T-cell surface glycoprotein CD8 beta chain; Immunoglobulin
homology
C:Keywords: glycoprotein
150609 Length: 207 February 11, 2000 15:51 Type: P Check: 8972

1 MARPMMLWL CLQLPFCIN LLSQTPGYI LTKTNNSPEI VCPMKSEHNG
51 VVYRWNGQR QHFEFLFSS PLGKATYGN ISOEFSTIRG TSSYHSYRLH
101 INRLHSDNG TYCCITIOS QLLIGTGQL DVVDVLPSP MSLLVPLTKK
151 PMCRKXKNA INKKGACTPM VWVPLAAGAL LLLSLIPII RRRYRLRRRL
201 WYRAHR

!!AA_SEQUENCE 1.0
P1:139516 - ribosomal protein S3 - Acholeplasma axanthum
C:Species: Acholeplasma axanthum
C:Date: 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 13-Aug-1999
C:Accession: 139516
R:Roth, K.T.; Harrison, N.; Sears, B.B.
Int. J. Syst. Bacteriol. 44, 119-124, 1994
A:Title: Phylogenetic relationships among members of the class Mollicutes
deduced from rps3 gene sequences.
A:Reference number: 139515; MUID:94169035
A:Accession: 139516
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-257 <RES>
A:Cross-references: GB:L22465; NID:9456689; PIDN:AAA21916.1; PID:9456690
C:Genetics:
A:Gene: rps3
C:Superfamily: Escherichia coli ribosomal protein S3
139516 Length: 257 February 11, 2000 15:51 Type: P Check: 7310

1 MGRKNPPIGF RVGVRDMS KMYADKIVP ALVXEDAVIR KFLNKNNA
51 AVSHVEIERL KELVKRRKRV ITHSGKPCV VIGFEATIMK ETIASLEKLT
101 KKEIVFNVE VRKEPVATL VAQSMAEOL NRASFSTK NCYAKSIKVR
151 ANGITLSOR LGRREMARTE GYSEGOVPLH TLRADVEYAT ARAQTITGL
201 GIKWIIYGE ILPGQSHEEL RKEROSSASS NHGGKRRPS RKGPRRSQED
251 AATEGN
!!AA_SEQUENCE 1.0
P1:577489 - ribosomal protein S8 - Synechocystis sp. (strain PCC 6803)
N:Alternate names: protein sll1809
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 13-Aug-1999
C:Accession: 577489
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Aamizu, E.; Nakamura, Y.;
Miyajima, N.; Hirosewa, M.; Sugitani, M.; Sasamoto, S.; Kimura, T.; Hosouchi,
T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimo, S.;
Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium
Synechocystis sp. PCC6803. II. Sequence determination of the entire genome and
assignment of potential protein-coding regions.
A:Reference number: 574322; MUID:97061201
A:Accession: 577489
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-133 <KAN>
A:Cross-references: EMBL:D90905; GB:AB001339; NID:91652360; PIDN:BA17336.1;
PID:d1018069; PID:91652414
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June
1996
C:Genetics:
A:Gene: rps8
C:Superfamily: Escherichia coli ribosomal protein S8
C:Keywords: protein biosynthesis; ribosome
577489 Length: 133 February 11, 2000 15:51 Type: P Check: 8926

1 MASTDITDM LTRIRNACV RHSTTOVPT KXTLSIAKVL KSEGFEDYS
51 ETEGINKML VTLTKNGKT ROLPINTLQR VSKPOLRVYS PSKRIPRVLS
101 GIGAIYVS HGIMTDRAR ROGIGGEIIC YIW
!!AA_SEQUENCE 1.0
P1:701947 - hypothetical protein F1104.1 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)

```

C>Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 13-Aug-1999  
 C:Accession: T01947  
 R:Abu-Threideh, J.; Stoneking, T.; Langston, Y.; Trevaaskis, E.  
 Submitted to the EMBL Data Library, October 1998  
 A:Description: The sequence of A. thaliana Fl104.  
 A:Reference number: 214466  
 A:Accession: T01947  
 A:Status: Preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-252 <ABU>  
 A:Cross-references: EMBL:AF096370; NID:g3695372; PIDN:AAC62776.1; PID:g3695375  
 C:Genetics:  
 A:Map position: 4  
 A:Note: Fl104.1  
 C:Superfamily: Lens fiber membrane major intrinsic protein

T01947 Length: 252 February 11, 2000 15:51 Type: P Check: 4063 ..

1 MPINRIAIGT PGEASRPDAI RAAPAEFFSM VIFVAGQS GMAYGKLTGD  
 51 GPATPAGLVA ASLSHAFALF VAASVGANVS GGHVNDAYTF GAFIGNTTL  
 101 LRAILWIAQ LGAVVACLL LKVGSGMET AAFSLSYGV PMNAVAFEIV  
 151 MTEGLVTVY ATAVDPKRGD IGIAPLAIG LIYGANILVG GAFDGSNMP  
 201 AVSFGPAVVS WITNMWYV VGPFGAIA AIYDTIFIG SNGHEPLPSN  
 251 DF

1AA\_SEQUENCE 1.0  
 P1:A39234 - opsin - bluebottle fly (Calliphora vicina)  
 C:Species: Calliphora vicina  
 C>Date: 22-Jan-1993 #sequence\_revision 22-Jan-1993 #text\_change 13-Aug-1999  
 C:Accession: A39234  
 R:Huber, A.; Smith, D.P.; Zuker, C.S.; Paulsen, R.  
 J. Biol. Chem. 265, 17906-17910, 1990  
 A:Title: Opsin of Calliphora peripheral photoreceptors R1-6. Homology with Drosophila Rh1 and posttranslational processing.  
 A:Reference number: A39234; MUID:91009258  
 A:Accession: A39234  
 A:Molecule type: mRNA  
 A:Residues: 1-371 <HUB>  
 A:Cross-references: GB:M58334; GB:J05596; NID:g156511; PIDN:AAA62725.1; PID:g156512  
 C:Superfamily: vertebrate rhodopsin  
 C:Keywords: chromoprotein; G protein-coupled receptor; glycoprotein; lipoprotein; membrane protein; photoreceptor; retinal  
 F:317/Binding site: retinal (Lys) (covalent) #status predicted  
 A39234 Length: 371 February 11, 2000 15:51 Type: P Check: 9791 ..

1 MERKSTFLIG PSFALTNCS VTDKVTDPMA HLHVHPWNOF PAMEKMAKMF  
 51 LAAYMIAT ISMCGNGVY YIFSTISLR TPANLVLINL AISDGMIT  
 101 NTPMGGINLF YETWVGLPM CDYIGGLGSA FCGSSILSMC MISLDRIVVI  
 151 VKGNAGOPMT IKLAIKKIAL IWMASITWL APVFGSRVY PEGNITSCGI  
 201 DYLERDMNPR SYLIFYSTIV YLPLFLICY SYWITIAVS AHEKAREOA  
 251 KKNVNSLRS SEDADKSAEG KLAQVALVTI SLWFAMWTFY TIINTLGLFK  
 301 YEGITPLNTI WGACFAKSA CYNPIYVIGS HPRYGIALKE KCPCCVFGKY  
 351 DDGKASDATS QATNNESEK A  
 1AA\_SEQUENCE 1.0  
 P1:S40691 - opsin rh1 - fruit fly (Drosophila pseudoobscura)  
 C:Species: Drosophila pseudoobscura  
 C>Date: 19-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 13-Aug-1999  
 C:Accession: S40691; S65546; S24605

R:Carulli, J.P.; Hartl, D.L.  
 Genetics 132, 193-204, 1992  
 A:Title: Variable rates of evolution among Drosophila opsin genes.  
 A:Reference number: S40691; MUID:93012921  
 A:Accession: S40691  
 A:Molecule type: DNA  
 A:Residues: 1-374 <CAR>  
 A:Cross-references: EMBL:X65877  
 R:Carulli, J.P.  
 Submitted to the EMBL Data Library, November 1992  
 A:Reference number: S65546  
 A:Accession: S65546  
 A:Molecule type: DNA  
 A:Residues: 1-173 'W', 175-374 <CAR>  
 A:Cross-references: EMBL:X65877; NID:g9077; PIDN:CAA46708.1; PID:g9078  
 C:Genetics:  
 A:Gene: FlyBase:Dpse/tnaE  
 A:Cross-references: FlyBase:FBgn0012733  
 A:Introns: 3/2; 191/2; 240/3; 333/2  
 C:Superfamily: vertebrate rhodopsin  
 C:Keywords: chromoprotein; lipoprotein; retinal  
 F:320/Binding site: retinal (Lys) (covalent) #status predicted  
 S40691 Length: 374 February 11, 2000 15:51 Type: P Check: 7181 ..

1 MDSFAAVATQ LGPOFAAPSN GSVDKVTDP MAHLISPYWD QFPADPIMA  
 51 KILTAIITII GMISWCGNGV VIYIFATRS LTPANLVI NLAISDFGIM  
 101 ITNTPMGGIN LFETWVGLP MMCDIYAGLG SAFGCSIMS MCMISLDRIQ  
 151 VIYKMGARBP MTIPALGKI AYICFMSTIM CCLAVFGMS RYVDEGNLTS  
 201 CGIDYLEROM NPSRLIYFS IFYIYIPFL ICYSWIFIA AVASHEKAMR  
 251 EQAKKNVKS LRSSSDADKS AEGKIAKVAL VTISWFMW TPLIVINCNG  
 301 LRFEBELTPL NTIMGACPAK SAACYNPIYV GISHPKYLA LKECPCCVF  
 351 GKYDCKSSSE AOSQATTSEA ESKA

1AA\_SEQUENCE 1.0  
 P1:JC4304 - orphan G-protein-coupled receptor - human  
 N:Alternate names: V28 protein  
 C:Species: Homo sapiens (man)  
 C>Date: 16-Nov-1995 #sequence\_revision 08-Feb-1996 #text\_change 20-Sep-1999  
 C:Accession: JC4304  
 R:Rapport, C.J.; Schweickart, V.L.; Eddy Jr., R.L.; Shows, T.B.; Gray, P.W.  
 Gene 163, 295-299, 1995  
 A:Title: The orphan G-protein-coupled receptor-encoding gene V28 is closely related to genes for chemokine receptors and is expressed in lymphoid and neural tissues.  
 A:Reference number: JC4304; MUID:96011651  
 A:Accession: JC4304  
 A:Molecule type: mRNA  
 A:Residues: 1-355 <RAP>  
 A:Cross-references: GB:U20350; NID:g665580; PID:g665581  
 A:Experimental source: peripheral blood mononuclear cell  
 C:Comment: This protein is a cell-surface receptor which recognizes extracellular signals and transduces those signals into an intracellular response.  
 C:Comment: This protein is a key regulator of many immune and homeostatic responses, and interacts between the nervous and immune systems.  
 C:Genetics:  
 A:Gene: V28  
 A:Map position: 3pter-p21  
 C:Superfamily: vertebrate rhodopsin  
 C:Keywords: G protein-coupled receptor; lymphokine; transmembrane protein  
 F:35-57/Domain: transmembrane #status predicted <TM1>  
 F:66-88/Domain: transmembrane #status predicted <TM2>  
 F:104-125/Domain: transmembrane #status predicted <TM3>  
 F:146-165/Domain: transmembrane #status predicted <TM4>  
 F:197-217/Domain: transmembrane #status predicted <TM5>

Mon Feb 14 08:07:32 2000

pir.cat

Page 91

F:230-254/Domain: transmembrane #status predicted <TM6>  
F:275-296/Domain: transmembrane #status predicted <TM7>

JC4304 Length: 355 February 11, 2000 15:51 Type: P Check: 2863

1 MOGFESTATE NFEYDGLAE CYIGDIYVFG TVFLSTFYSV IFAIGLVNL  
51 LVFVLTNSK KPKSVTDIYL LNLALSDLE VALPEWTHY LINEKGLHNA  
101 MCKFTTAPF IGFFGSIFFI TVISIDRYLA IYLANSMNN RTVOHGVTIS  
151 LGVMAALIV AAPQFMFTK KENECLGDP EVLGEIPVL RVNENFELGF  
201 LPLPLMSYC YFRITOTLFS CNHKKAKAI KLILVIVF FLEWPPYNN  
251 IFLETLKLYD FPPSCDMKRD LRALSVTET VAFSHCLNP LIAAFAGEKF  
301 RRYLHLYGK CLAVLCGRSV HVDSSSESO RSRHGVISS NRTYHNSDGL  
351 ALLLL

!!AA\_SEQUENCE 1.0  
P1:I58186 - hypothetical G-protein coupled receptor - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 26-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 26-Aug-1999  
C:Accession: I58186  
R:Harrison, J.K.; Barber, C.M.; Lynch, K.R.  
Neurosci. Lett. 169, 85-89, 1994  
A:Title: cDNA cloning of a G-protein-coupled receptor expressed in rat spinal  
cord and brain related to chemokine receptors.  
A:Reference number: I58186; MUID:94323113  
A:Accession: I58186  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-354 <RES>  
A:Cross-references: EMBL:004808; NID:9439860; PID:9439861  
C:Superfamily: vertebrate rhodopsin  
C:Keywords: G protein-coupled receptor

I58186 Length: 354 February 11, 2000 15:51 Type: P Check: 4478

1 MPTSPFELDL ENFEYDSAE ACYIGDIYAF GTFFLSIFYS LVTFGLVGN  
51 LNVVALTNS RKSKSTIDYI LNLALSDLE FVALTEFWTH YLISHEGLHN  
101 AMCKLTAPF FICFFGCIFF ITVISIDRYL AYLANSMMN NRTVOHGVTI  
151 SLGVMAAIL VASPOFMFTK RKNDECLGDY PEVLQIIMVY LRNSEVNLG  
201 FVLPLINSF CYFRIVRTLE SCKNKKARA IRLILVVVY FLEWTPYNI  
251 VIFLETLKLY NFFPSGCMR DIRMALSYTE TVAFSHCLNP PRYAFAGEK  
301 FRRYLRHLYN KCLAVLCGRD VHAGFSTESQ RSRKDSITLSS LTHYTSDEG  
351 SLLL

!!AA\_SEQUENCE 1.0  
F1:A29667 - pulmonary surfactant protein B - bovine (fragment)  
N:Alternate names: pulmonary surfactant 1-phe protein  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 16-Oct-1998  
C:Accession: A29667; S02317  
R:Olafson, R.W.; Rink, U.; Ktjelland, S.; Yu, S.H.; Chung, J.; Harding, P.G.R.;  
Possmayer, F.  
Biochem. Biophys. Res. Commun. 148, 1406-1411, 1987  
A:Title: Protein sequence analysis studies on the low molecular weight  
hydrophobic proteins associated with bovine pulmonary surfactant.  
A:Reference number: A29667  
A:Accession: A29667  
A:Molecule type: Protein  
A:Residues: 1-60 <OL>  
R:Yu, S.H.; Chung, W.; Olafson, R.W.; Harding, P.G.R.; Possmayer, F.

Biochem. Biophys. Acta 921, 437-448, 1987  
A:Title: Characterization of the small hydrophobic proteins associated with  
pulmonary surfactant.  
A:Reference number: S02317; MUID:88025156  
A:Accession: S02317  
A:Molecule type: Protein  
A:Residues: 1-60 <YU>

A29667 Length: 60 February 11, 2000 15:51 Type: P Check: 8980

1 FPIPIPCNL LRLIKKIQK VIPKGVLTMT VAVGVHVPV LVGGIIQOLV  
51 IEXSVILKTD

!!AA\_SEQUENCE 1.0  
P1:S2494 - RNA N-glycosidase (EC 3.2.2.22) alpha-luffin precursor - smooth  
loofah  
N:Alternate names: protein synthesis inhibitor; ribosome-inactivating protein  
luffin-A  
C:Species: Luffa cylindrica (smooth loofah)  
C:Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 20-Aug-1999  
C:Accession: S22494; S26390; JH0202; A32542  
R:Kataoka, J.; Habuka, N.; Miyano, M.; Masuta, C.; Koiwai, A.  
Plant Mol. Biol. 18, 1199-1202, 1992  
A:Title: Nucleotide sequence of cDNA encoding alpha-luffin, a  
ribosome-inactivating protein from Luffa cylindrica.  
A:Reference number: S22494; MUID:92288316  
A:Accession: S22494  
A:Molecule type: mRNA  
A:Residues: 1-277 <KAT>  
A:Cross-references: EMBL:X62371; NID:919145; PID:CAA44229.1; PID:919146  
R:Islam, M.R.; Nishida, H.; Funatsu, G.  
Agric. Biol. Chem. 54, 2967-2978, 1990  
A:Title: Complete amino acid sequence of luffin-a, a ribosome-inactivating  
protein from the seeds of sponge gourd (Luffa cylindrica).  
A:Reference number: S26390; MUID:91248471  
A:Accession: S26390  
A:Molecule type: protein  
A:Residues: 21-53, 'L', '55', 'I', '57-86', 'SOL', '90-154', 'L', '156-157', 'I', '159-173', 'L', '175-209', 'SL', '212-267' <ISD>  
R:Islam, M.R.; Nishida, H.; Funatsu, G.  
Agric. Biol. Chem. 54, 1343-1345, 1990  
A:Title: Complete amino acid sequence of luffin-a, a ribosome-inactivating  
protein from the seeds of Luffa cylindrica.  
A:Reference number: JH0202; MUID:91197482  
A:Accession: JH0202  
A:Molecule type: protein  
A:Residues: 21-53, 'L', '55', 'I', '57-86', 'SOL', '90-154', 'L', '156-157', 'I', '159-173', 'L', '175-209', 'SL', '212-267' <ISD>  
A:Experimental source: seed  
R:Ramakrishnan, S.; Enghild, J.J.; Bryant Jr., H.L.; Xu, F.J.  
Biochem. Biophys. Res. Commun. 160, 509-516, 1989  
A:Title: Characterization of a translation inhibitory protein from Luffa  
aegyptiaca.  
A:Reference number: A32542; MUID:89246493  
A:Accession: A32542  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 21-28, 'G', '30', 'X', '32-34', 'K', '36-40' <RAM>  
C:Function: hydrolyzes the N-glycosidic bond of a specific adenosine in 28S  
rRNA thereby inactivating the ribosome  
C:Superfamily: RNA N-glycosidase; RNA N-glycosidase homology  
C:Keywords: glycoprotein; glycosidase; hydrolyase; seed; toxin  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-267/Product: RNA N-glycosidase alpha-luffin #status experimental <MAT>  
F:24-262/Domain: RNA N-glycosidase homology <RNG>  
F:268-277/Domain: carboxyl terminal propeptide #status predicted <CTP>  
F:48/53/97/104/225/246/Binding site: carbonyl (Asn) (covalent) #status  
experimental

F:90.179.182/Active site: Tyr, Glu, Arg #status predicted

S22494 Length: 277 February 11, 2000 15:51 Type: P Check: 429 ..

1 MKRFTVLIA IFVASTVEA DVRFSLSGSS STSYSKFIDG LRRALPSNGT  
C:Species: Luffa cylindrica (smooth loofah)  
C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 20-Aug-1999  
C:Accession: S23519; MUID:92353400  
R:Kataoka, J.; Haiduka, N.; Miyano, M.; Masuta, C.; Koiwai, A.  
Plant Mol. Biol. 19, 887-889, 1992  
A:Title: Nucleotide sequence of cDNA encoding beta-luffin, another  
ribosome-inactivating protein from luffa cylindrica.  
A:Reference number: S23519; MUID:92353400  
A:Accession: S23519  
A:Molecule type: mRNA  
A:Residues: 1-278 <KAT>  
A:Cross-references: EMBL:X62372; NID:919149; PIDN:CAA44230.1; PID:919150  
C:Superfamily: RNA N-glycosidase; RNA N-glycosidase homology  
F:26-264/Domain: RNA N-glycosidase homology <RNG>

11AA\_SEQUENCE 1.0  
P1:S23519 - beta-luffin - smooth loofah  
C:Species: Luffa cylindrica (smooth loofah)  
C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 20-Aug-1999  
C:Accession: S23519; MUID:92353400  
R:Kataoka, J.; Haiduka, N.; Miyano, M.; Masuta, C.; Koiwai, A.  
Plant Mol. Biol. 19, 887-889, 1992  
A:Title: Nucleotide sequence of cDNA encoding beta-luffin, another  
ribosome-inactivating protein from luffa cylindrica.  
A:Reference number: S23519; MUID:92353400  
A:Accession: S23519  
A:Molecule type: mRNA  
A:Residues: 1-278 <KAT>  
A:Cross-references: EMBL:X62372; NID:919149; PIDN:CAA44230.1; PID:919150  
C:Superfamily: RNA N-glycosidase; RNA N-glycosidase homology  
F:26-264/Domain: RNA N-glycosidase homology <RNG>  
S23519 Length: 278 February 11, 2000 15:51 Type: P Check: 9157 ..

1 MNRFFFLSL ILIAFTVE GANVSFSLSG ADSKYSKFI TALKALPSK  
51 EVKSNIPILL PSASGASRYI LMGLSNYDAK AITMAIDVTN YIMGLVNS  
101 TSYFNEDSA KLASQYVFKG STIVLPYSG NYERLQNAAG KVAEKPIPLGF  
151 NAFDSATSL FHDSTAAAG ALFVITQTTA EASREKYEIG QITERIPKNE  
201 VSPALSLE NEWSALSQOI QLAQTNGAF RTPVITDKR GORVEIKDYN  
251 SKVVTNITKL LNKQNTAAAF DDGIPTKH  
11AA\_SEQUENCE 1.0  
P1:JN0108 - luffin-b - smooth loofah  
C:Species: Luffa cylindrica (smooth loofah)  
C:Date: 04-Sep-1998 #sequence\_revision 04-Sep-1998 #text\_change 07-May-1999  
C:Accession: JN0108  
R:Islam, M.R.; Hirayama, H.; Funatsu, G.  
Agric. Biol. Chem. 55, 229-238, 1991  
A:Title: Complete amino acid sequence of luffin-b, a ribosome-inactivating  
protein from sponge gourd (luffa cylindrica) seeds.  
A:Reference number: JN0108; MUID:91248488  
A:Accession: JN0108  
A:Status: Preliminary  
A:Molecule type: Protein  
A:Residues: 1-250 <ISL>  
C:Superfamily: RNA N-glycosidase; RNA N-glycosidase homology  
F:5-246/Domain: RNA N-glycosidase homology <RNG>

JN0108 Length: 250 February 11, 2000 15:51 Type: P Check: 7681 ..  
1 ANVSFSLSGA DSKYSKFI ALRKALPSKE KVSNIPLLP SASGASRYIL  
51 MOLSNYDAK IMAIDVTN YIMGLVNST SYFNESDAK LASQYFKGS  
101 TLVTPYSGN YERLQNAAG IREKIPILGR ALDSALTSIF HYDSTAAAA  
151 FLVITQTTAE ASREKYEIG IIRIPKNEY PSPALSLEN EAMSLSQOI

201 QLAQTNGAF RTPVITDKR GORVEITNLA SKVOIKDVS KLLNKQNTA

11AA\_SEQUENCE 1.0  
P1:S65052 - pistil-specific protein stsl4 precursor - potato  
C:Species: Solanum tuberosum (potato)  
C:Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 20-Aug-1999  
C:Accession: S65052  
R:van Elstik, G.J.; Wogens, M.; Rulter, R.K.; van Herpen, M.M.A.; Schrauwen,  
J.A.M.; Willems, G.J.  
Plant Mol. Biol. 30, 171-176, 1996  
A:Title: Molecular analysis of a pistil-specific gene expressed in the stigma  
and stilar cortex of Solanum tuberosum.  
A:Reference number: S65052; MUID:96197407  
A:Accession: S65052  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-214 <VAN>  
A:Cross-references: EMBL:X82652; NID:91236784; PIDN:CAA57976.1; PID:el24149;  
PID:91236785  
C:Genetics: stsl4  
C:Superfamily: pathogenesis-related leaf protein  
S65052 Length: 214 February 11, 2000 15:51 Type: P Check: 5285 ..

1 MEVLSTAMAC LVYIYIYD EKKRELKVR NKMTNLFQ FLTLTASSL  
51 THISAQTVPP PPEPTSAAT PPSRAOEFLL DAHNKARSEV GVGPLTWSM  
101 LAKETSLVR YQRDKONCFE ANLSNGRYG NQLMASGTVV TPRMAVDSMV  
151 AEKRYNEN NSCTGDDKCG VYQIYWKKS IELGCAQRTIC YEGPATLVC  
201 FYNPGNVIG EKPY

11AA\_SEQUENCE 1.0  
P1:JC4822 - acyl carrier protein - Bacillus subtilis  
N:Alternate names: 8.5K protein  
C:Species: Bacillus subtilis  
C:Date: 15-Aug-1996 #sequence\_revision 15-Oct-1996 #text\_change 20-Aug-1999  
C:Accession: JC4822; A69582  
R:Oguero, A.; Kakashita, H.; Takamatsu, H.; Nakamura, K.; Yamane, K.  
Gene 172, 17-24, 1996  
A:Title: The effect of Srb, a homologue of the mammalian SRP receptor  
alpha-subunit, on Bacillus subtilis growth and protein translocation.  
A:Reference number: JC4819; MUID:96257247  
A:Accession: JC4822  
A:Molecule type: DNA  
A:Residues: 1-77 <OGU>  
A:Cross-references: DDBJ:D64116; NID:91389548; PIDN:BAI0975.1; PID:dl011632;  
PID:9137013  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo,  
V.; Batters, M.G.; Besieres, P.; Bolochin, A.; Borchert, S.; Boriss, R.;  
Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;  
Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, S.;  
J.J.; Conerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denicoff, F.; Devine, K.M.;  
Duesterhoft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;  
Faber, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi,  
A.; Gallenro, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Goldthly, E.J.; Grand,  
G.; Guisepi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.;  
Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.;  
Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Bianchard, M.; Klein, C.;  
Kobayashi, Y.; Koester, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kuita, K.;  
Lapdus, A.; Lardinois, S.  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda,  
S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Mosati, D.;  
Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogasawa, K.; Ogilvara, A.; Oudega,  
B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Portwolk, S.; Prescott,  
A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds,  
S.; Rieger, M.; Rivolet, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato,



T.: Scanlon, E.  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serior, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpil, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenberg, M.; Vannier, F.; Vassart, A.; Vial, A.; Wambolt, R.; Weder, E.; Weder, H.; Wetzinger, T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A:Reference number: A69580; MUID:98044033  
A:Accession: A69582  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-77 <KUN>  
A:Cross-references: GB:299112; GB:AL009126; NID:92633902; PIDN:CAB13465.1; PID:el185183; PID:92633964  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: acpA; str  
C:Function:  
A:Description: carrier of the growing fatty acid chain; growing fatty acid chain is covalently bound to 4'-phosphopantetheine prosthetic group  
A:Pathway: fatty acid biosynthesis  
C:Superfamily: acyl carrier protein; acyl carrier protein homology  
C:Keywords: fatty acid biosynthesis; phosphopantetheine; phosphoprotein  
F:2/3/Domain: acyl carrier protein homology <ACP>  
F:3/Binding site: phosphopantetheine (Ser) (covalent) #status predicted  
JC4822 Length: 77 February 11, 2000 15:51 Type: P Check: 8148 ..

1 MADLERVTR IIVRLGVDE ADVKLEASF EDGANSULV VEIWELEDE  
51 FDMISDEDA EKATVGDV NVINQC  
11AA:SEQUENCE 1.0  
P1:B65003 histidine transport system permease protein HisQ - *Escherichia coli* (strain K-12)  
C:Species: *Escherichia coli*  
C>Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 20-Aug-1999  
C:Accession: B65003  
R:Baltner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y. Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of *Escherichia coli* K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: B65003  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-228 <BLAT>  
A:Cross-references: GB:AE000319; GB:U00096; NID:91788634; PIDN:AAC75368.1; PID:91788646; UMG:92308  
A:Experimental source: strain K-12, substrain MG1655  
C:Genetics:  
A:Gene: hisQ  
C:Superfamily: histidine permease protein M  
B65003 Length: 228 February 11, 2000 15:51 Type: P Check: 7651 ..

1 MUXGSSVIL QGALVLELA ISSVVLVII GLIGAGKLS QNRSLGLFE  
51 GYTLIRGVP DVLMLLIF GLQALNVT EAKGYOIDI DMVAGIITL  
101 GTIYAVITE TFGKFMAYV KGHIAATAF GTRGQVFR IMFPMATYA  
151 LFGINNMVY ILKSTALVSL LGLDEVYAT QLAGKSTWEP EYFAIVCVI  
201 YLVFTVSNQ VLFLERRIS VGVKRALD  
11AA:SEQUENCE 1.0  
P1:B36263 - oligopeptide transport system permease protein oppB - *Escherichia coli*  
C:Species: *Escherichia coli*  
C>Date: 18-Jan-1991 #sequence\_revision 30-Sep-1997 #text\_change 20-Aug-1999  
C:Accession: G64871; B36263  
R:Baltner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y. Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of *Escherichia coli* K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: G64871  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-306 <BLAT>  
A:Cross-references: GB:AE000223; GB:U00096; NID:91787496; PIDN:AAC74326.1; PID:91787497; UMG:91244  
A:Experimental source: strain K-12, substrain MG1655  
R:Kashiyagi, K.; Yamaguchi, Y.; Sakai, Y.; Kobayashi, H.; Igarashi, K. J. Biol. Chem. 265, 8387-8391, 1990  
A:Title: Identification of the polyamine-induced protein as a periplasmic oligopeptide binding protein.  
A:Reference number: A36263; MUID:90256748  
A:Accession: B36263  
A:Molecule type: DNA  
A:Residues: 1-16 <KAS>  
A:Cross-references: GB:U05433  
C:Genetics:  
A:Gene: oppB  
C:Function:  
A:Description: probably responsible for mediating passage of peptides across the cytoplasmic membrane  
C:Superfamily: oligopeptide permease protein oppB  
C:Keywords: binding protein; protein-dependent transport system; oligopeptide transport; transmembrane protein  
F:9-25/Domain: transmembrane #status predicted <TW1>  
F:105-121/Domain: transmembrane #status predicted <TW2>  
F:141-157/Domain: transmembrane #status predicted <TW3>  
F:173-189/Domain: transmembrane #status predicted <TW4>  
F:228-244/Domain: transmembrane #status predicted <TW5>  
F:273-289/Domain: transmembrane #status predicted <TW6>  
B36263 Length: 306 February 11, 2000 15:51 Type: P Check: 5147 ..

1 MKFLIRKCL EALPLTFLI TSIFMMLA PGSPFGER LPEVMANIE  
51 AKYHLNDPIM TQFSTYKQL AHGDFGPEFK YKDSVNDLV ASSFYSAKL  
101 GAATFLAVI LGVSAGVIAA LKONTRWDYT VMGLAMGV IPSVYAPDL  
151 VMFAIILDM LGGGWNNGA LKFMILPVA LSLAYIASIA RITRSMIEV  
201 LHSNFRIRAR AKGLPMRRI LRAHLKPLL PVLSYMGAP VGIITGSMTI  
251 ETIYGLPGIG QLFVNGALNR DYSVLSLTI LVGALTILFN AIVDLYAVI  
301 DPRIRY  
11AA:SEQUENCE 1.0  
P1:C71163 - probable oligopeptide transport system permease protein AppC - *Pyrococcus horikoshii*  
C:Species: *Pyrococcus horikoshii*  
C>Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 26-Aug-1999  
C:Accession: C71163  
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamaya, A.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ouchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H. DNA Res. 5, 55-76, 1998  
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium *Pyrococcus horikoshii* OT3.  
A:Reference number: A71000; MUID:98344137

A:Accession: C71163  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-304 <KAW>  
 A:Cross-references: GB:AP000002; NID:g3236129; PID:d1030535; PID:g3256909  
 A:Experimental source: strain O73  
 A:Note: this accession replaces an interim accession for a sequence replaced by GenBank  
 C:Genetics:  
 A:Gene: PH0504  
 C:Superfamily: oligopeptide permease protein oppb

C71163 Length: 304 February 11, 2000 15:51 Type: P Check: 4618 ..

1 MEYIKLAFNR NKKFGFGML IAFITFGIG PLTFPASDG LYEEVGKIK  
 51 IASYSAKTLP PMTRENITTY TGRKEVLIHT LGTDKCKDL YTELYGLRT  
 101 SLWVAFLAAT IGTITGITIG FVSGYKGLV DELLMFVNI MLVPSIVLL  
 151 ILVAAVLEAR SPEIQALITG LTNMPVARS VRAQTLSLKN REFMKSKIM  
 201 GVGDRLRIIVE DILPMNISYI FMTGILQVSG AILASATLDF IGLGPTTMS  
 251 LGVILQKAIM HNALQFGWMW WFIPPGLEIT LIITSLFFIN LGLEEVNPR  
 301 LRGV

11AA\_SEQUENCE 1.0  
 P1:564728 - protein secretion protein xcp - Pseudomonas putida  
 C:Species: Pseudomonas putida  
 C:Date: 06-Dec-1996 #sequence\_revision 06-Dec-1996 #text\_change 20-Aug-1999  
 C:Accession: S64728; S47505  
 R:de Groot, A.; Kijner, J.J.; Filloux, A.; Tommassen, J.  
 Mol. Gen. Genet. 250, 491-504, 1996  
 A:title: Characterization of type II protein secretion (xcp) genes in the plant growth-stimulating Pseudomonas putida, strain WCS358.  
 A:Reference number: S64724; MUID:96186881  
 A:Accession: S64728  
 A:Molecule type: DNA  
 A:Residues: 1-482 <DEG>  
 A:Cross-references: EMBL:X81085; NID:g531737; PIDN:CAAS6980.1; PID:g531741  
 C:Genetics:  
 A:Gene: xcpR  
 C:Superfamily: secretion protein xcpR

S64728 Length: 482 February 11, 2000 15:51 Type: P Check: 6546 ..

1 MLPYRLARQA GLAMPAGQG WQMLRRDAD SEVLQELNRY HQGPALAYL  
 51 EPAPFDQOLA QLYQGDNAI AALLEGIGDQ VDDLSLMSK PRIEDLEKSD  
 101 DEAPVILIN GLFGQALRLR ASDHIEFFE QSLVRLARD GHLREVLAP  
 151 RALSAMLYSR IKVMARLDA EKROPQDRI TLRAGREVD VRVSLPEIH  
 201 GERVYKVLUD KQSLIALDN LGMPAVLVHG LRSCLARPNG IYLSGTPTGS  
 251 GKTTIVLASL NSLINDSRNI LTVEDPEYA IAGIGGTAIN PRAGITFASG  
 301 LRLAIRDDPD VIMLGEIRDQ ETAQIAVQAS LTHGLVLSL HTNSAVGAVT  
 351 RLDDMGIEFP LIASCLRGVL AQRVLRRLCQ CAVAOPLQTA ERDLPELCA  
 401 LGSYHAYVC EOCQSGGYVG RLGYEFIEL DAGLIALYD GASEQAMQDY  
 451 LADRRSLVA MASDCLARGE TSLAEVLKVV QG

11AA\_SEQUENCE 1.0  
 P1:552163 - sucrose specific repressor - Escherichia coli  
 C:Species: Escherichia coli  
 C:Date: 07-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 26-Aug-1999  
 C:Accession: S52163

R:Bockmann, J.  
 submitted to the EMBL Data Library, September 1994  
 A:Reference number: S52160  
 A:Accession: S52163  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-331 <BOC>  
 A:Cross-references: EMBL:X81461; NID:g608705; PID:g608709  
 C:Superfamily: lac repressor

S52163 Length: 331 February 11, 2000 15:51 Type: P Check: 8969 ..

1 MASLKQVAVL AGVSMYTSR VMHNASVVR ATRDVLQAI QTLNYPVDS  
 51 ARKMRQGRK PSTLAVLAQD TATTPSVVDI LAIEQTAISE FGNNSEFINI  
 101 FSEDNARAA ROLAHRPDG IITTMGLRH ITPESLYGE NIYLANCVAD  
 151 DPALPSYIPD DYTAQVESTQ HLIAGYROP LCFWIPESAL ATGYRQFFE  
 201 QAWRDAGRDL AEVKQFNAT GDDHYTDLAS LNAHFKSGK PDEVDLIGN  
 251 DRAAFAYQV LIAKGVRIQ DVAVMGFDNL VGVGLFLPP LTTIQLPHDI  
 301 IGRERALHII EGREGRYTR IPCPLIRGS T

11AA\_SEQUENCE 1.0  
 P1:D70044 - transcription regulator GntR family homolog yvoA - Bacillus subtilis  
 C:Species: Bacillus subtilis  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Aug-1999  
 C:Accession: D70044  
 R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azavedo, J.; Barto, R.; Beckwith, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Busch, C.; Caldwell, B.; Capuano, N.J.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conerton, I.F.; Cummings, R.; Dagnan, R.A.; Denzot, F.; Devigne, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entlian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Galleron, N.; Ghim, S.Y.; Glaeser, P.; Goffeau, A.; Goldlighty, E.J.; Grandi, G.; Gillespie, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Kjaer, Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogikawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetalle, D.; Portoliz, S.; Prescott, A.M.; Prescann, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.  
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot, S.J.; Serron, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Yamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandendool, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Wetzenecker, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K.  
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A:title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
 A:Reference number: A69580; MUID:98044033  
 A:Accession: D70044  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-243 <KUN>  
 A:Cross-references: GB:Z99121; GB:AL009126; NID:g2635827; PIDN:CAB15508.1; PID:el186191; PID:g2636016  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: yvoA

C:Superfamily: transcription regulator GntR  
D70044 Length: 243 February 11, 2000 15:51 Type: P Check: 4428 ..

1 MNINKOSPIP IYYQIMEOLK TOIKNGELDP DMPLPSEREY AEQFGISKMT  
51 VRQALSNLVN EGIYRLKGR GTFVSKPME QALQGLTSTF EDMRSKMT  
101 GSRIDYOLI DSTEELAIL GCGHPSYIK ITRVRLANDI PMAIESHIP  
151 FELGELNES HFOSSYIDHI ERYNSIPISR AKOELPESA TTEANILGI  
201 OKGAPVLLIK RTTYLQNGTA FEHAKSVYRG DRYTFVHYMD RLS

11AA\_SEQUENCE 1.0  
P1:548826 - NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain PSTT precursor - wild cabbage  
C:Species: Brassica oleracea (wild cabbage)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 26-Aug-1999  
A:Accession: S48826  
R:Posson, B.J.; Downs, C.G.; Davies, K.M.; Morris, S.C.; Buchanan-Wollaston, V. submitted to the EMBL Data Library, September 1994  
A:Description: Nucleotide sequence of a cDNA clone from broccoli (Brassica oleracea L.) with high identity to the PSTT subunit of NADH: ubiquinone oxidoreductase.  
A:Reference number: S48826  
A:Accession: S48826  
A:Molecule type: mRNA  
A:Residues: 1-215 <PDB>  
A:Cross-references: EMBL:X82274; NID:9562281; PIDN:CA57725.1; PID:9562282  
C:Genetics:  
A:Gene: PSTT  
A:Genome: nuclear  
A:Superfamily: psbc protein  
C:Keywords: mitochondrion; NAD; oxidoreductase

S48826 Length: 215 February 11, 2000 15:51 Type: P Check: 9647 ..

1 MAMITNTAT RLPVLQSHR AAASHLHTS LPALSPATTP TSYTRBPSS  
51 TSAIPPGLSK TAEFVLSVD DLMNWRGRS IMPMTFGLAC CAVENMHTGA  
101 ARYDLDRCGI IFRSPRQSD CMIVAGLTIN KMAPLRKRY DQMPERAWYI  
151 SMSCANGCG YNHSYSVVR GCDRIYVDI YVPGCPFAE ALTYGLQLQ  
201 KTRNRKDFL HWMNK

11AA\_SEQUENCE 1.0  
P1:S75086 - Iron-stress chlorophyll-binding protein - Synechocystis sp. (strain PCC 6803)  
N:Alternate names: protein sl10247  
C:Species: Synechocystis sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 26-Aug-1999  
A:Accession: S75086  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirosewa, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naito, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.  
A:Reference number: S74322; MUID:97061201  
A:Accession: S75086  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-342 <RAN>  
A:Cross-references: EMBL:D90910; GB:AB001339; NID:g1652956; PIDN:BAAL7948.1; PID:dt018681; PID:g1653031  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:  
A:Gene: isiA  
C:Superfamily: photosystem II chlorophyll a-binding protein psbc  
S75086 Length: 342 February 11, 2000 15:51 Type: P Check: 4153 ..

1 MOTYNDTVQ YEMWAGNAR ADQSLFTA HVAQALTAFA WAGFTLFEI  
51 SRFDPQAMG DQGLILPHL AILGWVGDG GQIVDTYPRF VIGSIHLIAS  
101 AYLNGALFH TIRAPDLST LKQGRKRFH TWENPOOLGI ILGHILFLG  
151 AGALLIAGKA MWMGLYDAR TOTVALVSQP TLDPVIYVG QHFASSISL  
201 EDLVGCHTFV GFLIGGGIV HILVPLGMA KYVLFSGEA ILSYSLGIA  
251 LAGFYAAVFC AVNTLAPPE FYGPLAIKL GIPYFADYV ELPMAHHSR  
301 AMIANHFFL AEFLLQGLW HALRALGDFE KRYEQATDSL QT

11AA\_SEQUENCE 1.0  
P1:A64834 - probable fimbrial protein-like protein ycbQ precursor - Escherichia coli  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 26-Aug-1999  
A:Accession: A64834  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y. Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: A64834  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-182 <BLAT>  
A:Cross-references: GB:AE000196; GB:U00096; NID:g1787169; PIDN:AAC74024.1; PID:g1787170; UMG:p0938  
A:Experimental source: strain K-12, substrain MG1655  
C:Genetics:  
A:Gene: ycbQ  
C:Superfamily: type 1 fimbrial protein  
C:Keywords: fimbria  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-182/Product: probable fimbrial protein-like protein ycbQ #status predicted <MAT>

A64834 Length: 182 February 11, 2000 15:51 Type: P Check: 3224 ..

1 MIEKRSVLT AFITVCATS SYNAADDNAI TDGVTENGK VIAPACTIVA  
51 ATKDSVYTLR DVSATKLTQN GOVSGVQIDV PIELKDDTT VTKNAFTFN  
101 GTADTGTITA FANOASSDAA TNVALQWYMN DGTITATPDT ETGNILLQDG  
151 DQITFEVYD IATGATSGN VNATNHNH YX

11AA\_SEQUENCE 1.0  
P1:D71484 - probable ribosome releasing factor - Chlamydia trachomatis (serotype D strain W3/Cx)  
C:Species: Chlamydia trachomatis  
C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 26-Aug-1999  
A:Accession: D71484  
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, W.F.; Dlugosz, E.; Tatusov, R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W. Science 282, 754-759, 1998  
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis.  
A:Reference number: A71570; MUID:9900809  
A:Accession: D71484  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-179 <ARN>

A:Cross-references: GB:AE001338; GB:AE001273; NID:g3329126; PID:g3329129  
 A:Experimental source: serotype D, strain UW-3/Cx  
 C:Genetics:  
 A:Gene: ftr  
 C:Superfamily: ribosome releasing factor  
 D1484 Length: 179 February 11, 2000 15:51 Type: P Check: 1871 ..

1 MTLASAKEM VGVLFPEKE TRGFRGKAH PALVEVAYE YGVTMLSD  
 51 IASISVSDMR QLLSPYDAG TVSAISKGL AANLNGPIV EGAVRLNVP  
 101 EPREEFRREV IKQLKRSEE AKVAIRNIR TENDRLKDD NLTEDAVKSL  
 151 EKKIQELTRK FCKQIEELAK OKAEALATV

!!AA\_SEQUENCE 1.0  
 P1:S74877 - phosphate transport system permease protein C.2, 35K -  
 Synecocystis sp. (strain PCC 6803)  
 N:Alternate names: protein sir1248  
 C:Species: Synecocystis sp.  
 A:Variety: PCC 6803  
 C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 26-Aug-1999  
 C:Accession: S74877  
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.;  
 Miyajima, N.; Hikosawa, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi,  
 T.; Matsumo, A.; Muraaki, A.; Nakazaki, N.; Nario, K.; Okumura, S.; Shimpo, S.;  
 Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.  
 DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium  
 Synecocystis sp. PCC6803. II. Sequence determination of the entire genome and  
 assignment of potential protein-coding regions.  
 A:Reference number: S74322; MUID:97061201  
 A:Accession: S74877  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-328 <RAN>  
 A:Cross-references: EMBL:D90909; GB:AB001339; NID:g1652844; PIDN:BAAL7838.1;  
 PID:01018571; PID:g1652920  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June  
 1996  
 C:Genetics:  
 A:Gene: pstC-2  
 C:Superfamily: phoW protein  
 C:Keywords: phosphate transport; transmembrane protein  
 S74877 Length: 328 February 11, 2000 15:51 Type: P Check: 7310 ..

1 MVEGFRSE FTIARGDIT QGNSLTGILD WGFQQLTRLC SLGVVILGW  
 51 IANVETDAR PAIAFGWEF IVSQWDSG QLEGGPIYF GSVVSSIAL  
 101 IFALPIGLAV AVMTSENLP APYAVPIAFV VELASIPSV IIGMGIVF  
 151 IPIVMVQGA LEKYIGLPI FGTEPEPSM LVAGVLVTM IIPITASIR  
 201 DILISVPSI RSASMALGAT RWETICVIL PSASSCIITA TITLAGRALG  
 251 ETMAVTWVG NSNITASL APGYTIPSV ANQFAEAVDE LHIGALMTLA  
 301 LILEVITLGI NSLAVLMVAT IRRGESH

!!AA\_SEQUENCE 1.0  
 F1:A46259 - recombinant protein reca homolog - Arabidopsis thaliana (fragment)  
 N:Alternate names: recombinase A  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 26-Aug-1999  
 C:Accession: A46259  
 R:Centilli, H.; Osman, M.; Grandoni, P.; Jagendorf, A.T.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 8068-8072, 1992  
 A:Title: A homolog of Escherichia coli RecA protein in plastids of higher  
 plants  
 A:Reference number: A46259; MUID:92390390

A:Accession: A46259  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-438 <CER>  
 A:Cross-references: GB:M98039; NID:g166840; PIDN:AAA32855.1; PID:g166841  
 A:Note: sequence extracted from NCBI backbone (NCBIN:112803, NCBI:P:112805)  
 C:Superfamily: recombinant protein reca  
 C:Keywords: ATP; DNA binding; DNA recombination; DNA repair; P-loop; SOS  
 response  
 F:141-148/Region: nucleotide-binding motif A (P-loop)  
 F:215-220/Region: nucleotide-binding motif B  
 F:147/Binding site: ATP (lys) #status predicted  
 A46259 Length: 438 February 11, 2000 15:51 Type: P Check: 9097 ..

1 DSQVLVSLKL NPSFTPLSPL FPFPCSSFS PSRLSSCYG RLKSPVTIV  
 51 AAKKLSHKIS SEFDDRINGA LSPDADSRL DRQKLEAM NDINSEFGK  
 101 SVTRLGAGG ALVETSSGI LTIDLAGGG LPKGRVETI GPSSGKTTL  
 151 ALHAIAYOK LGNAMLVDA EHAEDPAYK ALGVENILI VQCPDNGEMA  
 201 LETADMRCS GAVDLICVDS VSALTFRAT EGEIGMOQG LQARIMSOAL  
 251 RKMSGNASKA GCTLIFLNOI RYKIGVYGN PEVTSGIAL KFEASVLEI  
 301 RSKGRKSKS GDEDIGLAR VRYQSKVSR PYQAEFEIM FGEGVSKLGC  
 351 VLDCATIMEV VYKGSWYSY EDQRLGQRE KALQHLREMP ALQDEIEKTV  
 401 RLMLDGEVH RSTPLMSSS SSASHREDE EDSLDDFO

!!AA\_SEQUENCE 1.0  
 P1:J00661 - impB protein - Salmonella typhimurium plasmid TP110  
 C:Species: Salmonella typhimurium  
 C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 26-Aug-1999  
 C:Accession: J00661  
 R:Jodalyk, D.; Owen, D.; Strike, P.  
 Nucleic Acids Res. 18, 5045-5050, 1990  
 A:Title: DNA sequence analysis of the impB UV protection and mutation operon of  
 the plasmid TP110: identification of a third gene.  
 A:Reference number: J00659; MUID:90384799  
 A:Contents: plasmid TP110  
 A:Accession: J00661  
 A:Molecule type: DNA  
 A:Residues: 1-424 <LOD>  
 A:Cross-references: EMBL:X53528; NID:g47745; PIDN:CAA37608.1; PID:g47748  
 C:Genetics:  
 A:Gene: impB  
 A:Gene: plasmid TP110  
 C:Function:  
 A:Description: involved in UV protection and mutation  
 C:Superfamily: umuc protein  
 C:Keywords: DNA repair; induced mutagenesis; SOS mutagenesis  
 J00661 Length: 424 February 11, 2000 15:51 Type: P Check: 6434 ..

1 MFALADINSE YASCEKFRP DLREPEYIVL SNNDCCVIAR SPEAKALGIR  
 51 MGQPFQVBO MLEKKIHF SSVYALYHSM SQRVAVLES LSPAVEYSI  
 101 DEMFIDLRI NHCISPEFFG HQLEQVSW TGLTGVGIA PTKRLASAO  
 151 WATQWPOFS GVALTAENR NRIKLGLIO PVGEWVGWH RLTEKLALG  
 201 INTALQAOA NTAIFRNFS VILERTVREL NGESCISLEE APPAKOOIVC  
 251 SRSFGRITD KDAHQAVYO YAEPAEKLK GERQCRQVT TFVTSPEAV  
 301 KERCYSNAV EKLPLPTQDS RDIIAACA R LNHVWREGVR YMKAGVWLAD  
 351 FTSSGIAQPG LFDLQPRKN SEKLMTLDE LMQSGKAYW FAGGTAPFW

401 QMKREMLSOC YTTKMDIPL ARLG

11AA\_SEQUENCE 1.0  
 PI:F71946 flagellar motor switch protein - Helicobacter pylori (strain J99)  
 C:Species: Helicobacter pylori  
 C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 26-Aug-1999  
 A:Accession: F71946  
 R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonghe, B.L.; Carnell, G.; Tummino, P.J.; Caruso, A.; Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.E.; Trust, T.J.  
 Nature 397, 176-180, 1999  
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori.  
 A:Reference number: A71800; MUID:99120557  
 A:Accession: F71946  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-343 <ARN>  
 A:Cross-references: GB:AE001468; GB:AE001439; NID:g4154838; PIDN:AD05900.1; PID:g4154847  
 A:Experimental source: strain J99  
 C:Genetics:  
 A:Gene: flig  
 C:Superfamily: flagellar switch protein flig  
 F71946 Length: 343 February 11, 2000 15:51 Type: P Check: 9168 ..

1 MATLITRKQK AQDELMSSE KAILLIQV EDTGELIR LDIDSEITS  
 51 KOIYQNGCT KQGAVALSE FFAFQSNQY INTGLEYAR ELTTTLESE  
 101 EARKVMDKT KSLQTKNFA YLGKRPQOL ADLIIHQ TALILAME  
 151 APNAETLSY FPDEMAEIS IRMANLEIS PQYKRVSY LENTRESLTS  
 201 YKIEVGGLRA VAEIFNRLO KSATTLARI ESDVNLKLA IEMFTED  
 251 IAKIDNFAIR ELKVADKND LSLAIKISTQ DLTDFLNMA SRRAAEVE  
 301 EMQYLGAVKI KDYVAQRKI IEIVSLOEK GVIQTGEED VIE

11AA\_SEQUENCE 1.0  
 PI:A37763 - virD1 protein - Agrobacterium tumefaciens plasmid pTiC58  
 C:Species: Agrobacterium tumefaciens  
 C:Date: 31-May-1991 #sequence\_revision 31-May-1991 #text\_change 26-Aug-1999  
 A:Accession: A37763; S11838; A22666  
 R:Wang, K.; Herrera-Estrella, A.; Van Montagu, M.  
 J. Bacteriol. 172, 4432-4440, 1990  
 A:Title: Overexpression of virD1 and virD2 genes in Agrobacterium tumefaciens enhances T-complex formation and plant transformation.  
 A:Reference number: A37763; MUID:90330550  
 A:Accession: A37763  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-147 <MAN>  
 A:Cross-references: GB:M33673; NID:g142284; PIDN:AAA2210.1; PID:g142285  
 R:Rogovsky, P.M.; Powell, B.S.; Shirsu, K.; Lin, T.S.; Morel, P.; Zyprian, E.M.; Steck, T.R.; Kado, C.I.  
 Plasmid 23, 85-106, 1990  
 A:Title: Molecular characterization of the vir regulon of Agrobacterium tumefaciens: complete nucleotide sequence and gene organization of the 28.93-kbp region cloned as a single unit.  
 A:Reference number: S11838; MUID:90301800  
 A:Accession: S11838  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-147 <ROC>  
 A:Cross-references: EMBL:J03320; NID:g154781; PIDN:AAA91603.1; PID:g154795  
 R:Hagiya, M.; Close, T.J.; Tait, R.C.; Kado, C.I.  
 Proc. Natl. Acad. Sci. U.S.A. 82, 2669-2673, 1985

A:Title: Identification of pTiC58 plasmid-encoded proteins for virulence in Agrobacterium tumefaciens.  
 A:Reference number: A94037; MUID:85190558  
 A:Accession: A22666  
 A:Molecule type: DNA  
 A:Residues: 1-57, 'P', '59-105', 'SPAS', 'HAG'  
 A:Note: the authors translated the codon ATA for residue as Met, ATA for residue 79 as Met and ATA for residue 83 as Met  
 C:Genetics:  
 A:Gene: virD1  
 A:Genome: plasmid  
 C:Superfamily: virD1 protein  
 A37763 Length: 147 February 11, 2000 15:51 Type: P Check: 2312 ..

1 MGGSRPSS DIANORECV KYEGKRVST RLRSARESF SNOARLGIS  
 51 DSMAIRAVR RIGGLEIDA ETRHREAIL QSIGTLSSNI AALLSYAEN  
 101 PTMDLEALRA ERIAFKSPA DLDGLRSTL SVSRRIDGC SLIKDAL

11AA\_SEQUENCE 1.0  
 PI:A25063 - hypothetical virD1 protein - Agrobacterium tumefaciens plasmids  
 C:Species: Agrobacterium tumefaciens  
 C:Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 26-Aug-1999  
 A:Accession: A29826; A25063  
 R:Jayaswal, R.K.; Veluthambi, K.; Gelvin, S.B.; Slightom, J.L.  
 J. Bacteriol. 169, 5035-5045, 1987  
 A:Title: Double-stranded cleavage of T-DNA and generation of single-stranded T-DNA molecules in Escherichia coli by a virD-encoded border-specific endonuclease from Agrobacterium tumefaciens.  
 A:Reference number: A91846; MUID:88032822  
 A:Accession: A29826  
 A:Molecule type: DNA  
 A:Residues: 1-147 <JAY>  
 A:Cross-references: EMBL:M17989; NID:g142288; PIDN:AAA2211.1; PID:g142289  
 A:Note: plasmid pTiA6  
 R:Yanofsky, M.F.; Porter, S.G.; Young, C.; Albright, L.M.; Gordon, M.P.; Nestor, E.W.  
 Cell 47, 471-477, 1986  
 A:Title: The virD operon of Agrobacterium tumefaciens encodes a site-specific endonuclease.  
 A:Reference number: A90886; MUID:87028239  
 A:Accession: A25063  
 A:Molecule type: DNA  
 A:Residues: 1-147 <JAY>  
 A:Note: plasmid pTiA6NC  
 C:Genetics:  
 A:Gene: virD1  
 A:Genome: plasmid  
 C:Superfamily: virD1 protein  
 C:Keywords: endonuclease  
 A25063 Length: 147 February 11, 2000 15:51 Type: P Check: 4145 ..

1 MSKHTVTS ETAINQNSL NVEGFRVSA RLRSARETF SYQARLLGIS  
 51 DSMAIRAVR RIGGLEIDA HTRKEMAIL QSIGTLSSNV SMDLSAYED  
 101 PRSDLEAVRD ERIAFGEAFA ALDGLRSTL SVSRRIDGC SLIKGAL

11AA\_SEQUENCE 1.0  
 PI:S06883 - virD1 protein - Agrobacterium rhizogenes plasmid pTiA4b  
 C:Species: Agrobacterium rhizogenes  
 C:Date: 07-Apr-1994 #sequence\_revision 07-Apr-1994 #text\_change 26-Aug-1999  
 A:Accession: S06883  
 R:Hirayama, T.; Muranaka, T.; Okawa, H.; Oka, A.  
 Mol. Gen. Genet. 213, 229-237, 1988  
 A:Title: Organization and characterization of the virD genes from Agrobacterium rhizogenes.  
 A:Reference number: S06883; MUID:89039712  
 A:Accession: S06883  
 A:Molecule type: DNA

A:Residues: 1-147 <HIR>  
A:Cross-references: EMBL:X12867; NID:g38995; PIDN:CAA31350.1; PID:g38999  
C:Genetics:  
A:Gene: vird1  
A:Genome: plasmid  
C:Superfamily: vird1 protein  
S06883 Length: 147 February 11, 2000 15:51 Type: P Check: 2761 ..

1 MSGSKRTSS DAINQRVGA TVEGFRVST RLRSAYESF SHQARLGLS  
51 DSAIRAVR RIGGFLEIDA ETRKMEIL LSTGTLSSNI AALLSAVEN  
101 PTMDLELRA ERIRAGEFRA DLGLKLSIL SVSRRLDGC SMLKDSL

!!A\_SEQUENCE 1.0  
P1:S36564 - E2 protein - human papillomavirus type 45  
C:Species: human papillomavirus type 45  
C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 26-Aug-1999  
C:Accession: S36564  
R:Delius, H.; Hofmann, B.  
Submitted to the EMBL Data Library, August 1993  
A:Description: Primer-directed sequencing of human papillomavirus types.  
A:Reference number: S36469  
A:Accession: S36564  
A:Molecule type: DNA  
A:Residues: 1-368 <DEL>  
A:Cross-references: EMBL:X74479; NID:g397022; PIDN:CAA52576.1; PID:g397026  
C:Superfamily: papillomavirus E2 protein  
C:Keywords: DNA binding; early protein; transcription regulation  
S36564 Length: 368 February 11, 2000 15:51 Type: P Check: 5628 ..

1 MKKQTPRESL SERLSLQDK ILDRYENDSK DINSQISTYMO LIRLENALIF  
51 TAREHGITKL NHOVVPINI SKSKRAKAE LQNALKGLAO SKYNNEMTL  
101 QDTCEELMNT EPSQCFKGG KTVHYEDGN KDNQNMYYW DSIYYITETG  
151 IMDKTACVS YWGVYYIKDG DTYVYQFS ECEKYGNSNT WEVOYQGANVI  
201 DCNDSMCSST DDTVSATQIV RQLOHASTST PRTASVGTFR PHIQTPATKR  
251 PROGLTEQH HGRVNTVHN PLLCSTSNM KRRKVCSGNT PPIHLKGDK  
301 NSLKCLNRYL RKYADHYSEI SSTWHTGNCN KNTGILLTVY NSEVOGNFEL  
351 DVTIRPSVQ ISVGKMTI

!!A\_SEQUENCE 1.0  
P1:S36576 - E2 protein - human papillomavirus type 52  
C:Species: human papillomavirus type 52  
C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 26-Aug-1999  
C:Accession: S36576  
R:Delius, H.; Hofmann, B.  
Submitted to the EMBL Data Library, August 1993  
A:Description: Primer-directed sequencing of human papillomavirus types.  
A:Reference number: S36469  
A:Accession: S36576  
A:Molecule type: DNA  
A:Residues: 1-368 <DEL>  
A:Cross-references: EMBL:X74481; NID:g397038; PIDN:CAA52588.1; PID:g397042  
C:Superfamily: papillomavirus E2 protein  
C:Keywords: DNA binding; early protein; transcription regulation  
S36576 Length: 368 February 11, 2000 15:51 Type: P Check: 4491 ..

1 MESIPARLNA VOBKILDLYE ADSNDLNAQI EHKILIRMEC VLFYAKELG  
51 ITHIGHQVVP PMAVSKAKAC QAIELQALE ALKKTQYSTD GWTLOQTSLE  
101 MMRAPQKTF KKHGTYITVQ YDNCKNNTMD YTNMKETIYL GECECTIVEG

151 QVDYIGLYW CDGEKIYFVK FSNDAKOYV TGWEVHNGG QVIVCAPSVS  
201 SNEVSTTERA VHLCTETSKT SAVSVGAKDT HLOPPKRRR PDVDSRNTK  
251 YPNNLRGQO SYDSTTRGLV TATECTNKR VAHTCTAPI IHLKDPNSL  
301 KCLRIRYKTH KSLYQOIST WHMTSNECTN NKLGIYTVY SDETROOFL  
351 KTVKIPYVQ VIOGVMSL

!!A\_SEQUENCE 1.0  
P1:S36552 - E2 protein - human papillomavirus type 3  
C:Species: human papillomavirus type 3  
C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 26-Aug-1999  
C:Accession: S36552  
R:Delius, H.; Hofmann, B.  
Submitted to the EMBL Data Library, August 1993  
A:Description: Primer-directed sequencing of human papillomavirus types.  
A:Reference number: S36469  
A:Accession: S36552  
A:Molecule type: DNA  
A:Residues: 1-383 <DEL>  
A:Cross-references: EMBL:X74462; NID:g397005; PIDN:CAA52472.1; PID:g397009  
C:Superfamily: papillomavirus E2 protein  
C:Keywords: DNA binding; early protein; transcription regulation  
S36552 Length: 383 February 11, 2000 15:51 Type: P Check: 4819 ..

1 METLANRLDV CODKILELYE KSDSKLEDOJ MRMQLRLBQ ALTYKARECG  
51 LTHIGHQVVP PLSYTKAKAR SAIEVHYSLO QLOHSHADQ PWTLRDTSRE  
101 MMDTVPRKCM KRGGLTVEFR YDGENKAMC YQWRELIYO NYTDNNMKV  
151 AGLVSHGELY YMBEQKTFY VKFKODARVY GDTGWDVAV GGKYLHDSF  
201 DPVSTREIP APGLYACTT QAPLOAOVGA SEGEQKORR LETYVGEQOQ  
251 QOQOQOQOQOQ HTQTPAQOTT ERAHQPLDIT RTDRDITCP HPIGHRSDDP  
301 CVPYHLRGD PNLGCFRFR LNKGNKLYS RSTSTWRWSC ESENOCAVYT  
351 IWTSYGORE AFLSTVKVP GIVYLGHMS MFT

!!A\_SEQUENCE 1.0  
P1:S36535 - E2 protein - human papillomavirus type 10  
C:Species: human papillomavirus type 10  
C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 26-Aug-1999  
C:Accession: S36535  
R:Delius, H.; Hofmann, B.  
Submitted to the EMBL Data Library, August 1993  
A:Description: Primer-directed sequencing of human papillomavirus types.  
A:Reference number: S36469  
A:Accession: S36535  
A:Molecule type: DNA  
A:Residues: 1-376 <DEL>  
A:Cross-references: EMBL:X74465; NID:g396901; PIDN:CAA52492.1; PID:g396905  
C:Superfamily: papillomavirus E2 protein  
C:Keywords: DNA binding; early protein; transcription regulation  
S36535 Length: 376 February 11, 2000 15:51 Type: P Check: 273 ..

1 METLANRLDA CODKMLELYE KSDSKLEDOJ TMHHLRIVEN ALTYKARECG  
51 LTHIGHQVVP PLSYTKAKAR NAIIVHALQ QLOESAYANE PWTLRDTSRE  
101 MMDTAPKGCW KRGGLTVEFR YDGENKAMC YQWRELIYO NYSDDRMKV  
151 PGKYSIEGLY YTHENNNIY VNFKODACYV GETGKREIVY GGKYLHDAF  
201 DPVSTREIS TPGPYCTSNV TPASTQAVG ASEGEQKQO RUEAVDGOHQ  
251 QORQSKDST OKAERBAGQ VSDRRLDCL TRSAHPVRHP SDPCAPYIH

```

301  LAGDENSILKC FRYRLHNGKR KLYSRSSSTW RMCSESENQA AFVTLMTSD
351  TORTEFLNV KVPPGIQVIL GYMSIF

!!A:SEQUENCE 1.0
P1:A44215 - EUS1 protein - equine herpesvirus 1 (strain Kentucky A)
C:Species: equine herpesvirus 1
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999
C:Accession: A44215
R:Breedem, C.A.; Yalamanchili, R.R.; Colle III, C.F.; O'Callaghan, D.J.
Virology 191, 649-660, 1992
A:Title: Identification and transcriptional mapping of genes encoded at the
IR/US junction of equine herpesvirus type 1.
A:Reference number: A44215; MUID:93079867
A:Accession: A44215
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-303 <BRE>
A:Cross-references: GB:M80429
C:Superfamily: herpesvirus US2 protein

A44215 Length: 303 February 11, 2000 15:51 Type: P Check: 5996

1  MGVLLITVVT VDHRKALPN SSIDVDGHLW EFLSQCFVL ASEPLGIPIY
51  VRSADLRFSS SLLTLPRKC RPIVTRGDT AIALDRNGV YHEDRMGVSI
101  EMLSVLSGYN HLNSSLINQ PYHLVVGAA DICKPVFDLI PGPFRVYAE
151  IADFFHSMQ PPFVCGKLE TTPWTVEHN HPLKLAAG EDIVVGECHF
201  SKHSNSLVH PPTVNVITA VDDPARLEI PAFGRPLPRP RPSGGRAP
251  RRRSAPAPA RSTAAATPP RPDPRAPAA RRAQDTWME RLMGVGTGT
301  STR

!!A:SEQUENCE 1.0
P1:F43674 - US7 protein - human herpesvirus 2 (strain HG52)
C:Species: human herpesvirus 2
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 26-Aug-1999
C:Accession: F43674
R:McGeoch, D.J.; Moss, H.W.M.; McNab, D.; Frame, M.C.
J. Gen. Virol. 68, 19-38, 1987
A:Title: DNA sequence and genetic content of the HindIII 1 region in the short
unique component of the herpes simplex virus type 2 genome: identification of
the gene encoding glycoprotein g, and evolutionary comparisons.
A:Reference number: A43674; MUID:87111457
A:Accession: F43674
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-372 <MCG>
A:Cross-references: EMBL:X04798; NID:959900; PIDD:CAA28485.1; PID:959906
C:Superfamily: herpesvirus US7 protein

F43674 Length: 372 February 11, 2000 15:51 Type: P Check: 501

1  MPERSLQGLA ILGLWCATG LVRGPTVSL VDSLVDAGA VDGQVEED
51  LRVFGLHEV GAQVPHTVY DGLLEFHP LGNHCPRYVH VTLTACPRR
101  PAVAFILCRS THHASPAPV TLEGLARCP LKRVATARD YAGIYLRYW
151  VGSATNSRF VLGVALSNG TFVYNGSDY SCDPALPFS APRLGSSVY
201  TPGASRTPP RTTTPSSPR DPTAPGDTG TPAPASGEIA PPNSTRASAE
251  SRHLUYAIV IQAIAPASII AFVLSGSCIC FTHRCORRYR RPRGQIYNPG
301  GVSACVNDAA MARLGAELRS HPNTPPKPR RSSSSTWPS LTIASESEP
351  GPVVLVSVP RPSGPTAQ EV

!!A:SEQUENCE 1.0
P1:S58346 - coat protein V1 - tomato yellow leaf curl virus
C:Species: tomato yellow leaf curl virus
C>Date: 14-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 20-Sep-1999
C:Accession: S58346
R:Hong, Y.; Harrison, B.D.
Submitted to: The EMBL Data Library, February 1995
A:Description: Nucleotide sequences from tomato leaf curl viruses from
different countries: evidence for three geographically separate branches in
evolution of the coat protein of whitefly-transmitted geminiviruses.
A:Reference number: S58346
A:Accession: S58346
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-256 <HON>
A:Cross-references: EMBL:Z48182; NID:9944838; PIDD:CAA88227.1; PID:9944839
C:Superfamily: cassava latent virus coat protein

S58346 Length: 256 February 11, 2000 15:51 Type: P Check: 2806

1  MSRRPADMTI GGPVSKYRRL LSSISPYSKR AAVRTVGTGK GKEMANRPVN
51  RKPWFYRMR GPDYRGCEG PCKVQSFSR HDIHLGKM CISDVTGTC
101  LTHRGRKFC VASVYVLGKI WMDENIKTN HTNSVMEFLV RDRRPVDFQ
151  DFGDVEMED NEPTATYKN MHRDROYLR KWHATVTGGQ YASKEQALVK
201  EFVKVNNYV YNOQAGKYE NHSENLMLY MACTHASNPV YATLKIRIYF
251  YDSVTN

!!A:SEQUENCE 1.0
P1:J01248 - hypothetical 12K protein - chrysanthemum virus B
C:Species: chrysanthemum virus B
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Feb-1997
C:Accession: J01248
R:Deray, K.; Zaytsev, S.
J. Gen. Virol. 72, 2333-2337, 1991
A:Title: Nucleotide sequence and gene organization of the 3'-terminal region of
chrysanthemum virus B genomic RNA.
A:Reference number: J01248; MUID:92013948
A:Accession: J01248
A:Molecule type: genomic RNA
A:Residues: 1-106 <LEV>
A:Note: the authors translated the codon ATC for residue 57 as Tyr
C:Superfamily: barley stripe mosaic virus 14K protein
C:Keywords: transmembrane protein

J01248 Length: 106 February 11, 2000 15:51 Type: P Check: 566

1  MFLTPPPDHT KVLVAAGL SIVASILYIS RNTLPVGGH SHLLPHGCV
51  KDGTKIYVG GPRKLSLEG GENLPVOPWF LVILLSAIF LLSCHSGHRR
101  VCGQCH

!!A:SEQUENCE 1.0
P1:S41285 - coat protein - sweet potato G virus (fragment)
C:Species: sweet potato G virus
C>Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 20-Sep-1999
C:Accession: S41285
R:Colinet, D.; Kummer, J.; Lepoint, P.
Submitted to: The EMBL Data Library, December 1993
A:Reference number: S41285
A:Accession: S41285
A:Molecule type: DNA
A:Residues: 1-335 <COU>
A:Cross-references: EMBL:X76944; NID:9439687; PID:9439688
C:Superfamily: tobacco etch virus genome polypeptide; DEAD/H box helicase
homology
C:Keywords: coat protein

```

541285 Length: 355 February 11, 2000 15:51 Type: P Check: 3148 ..

1 SAEIYDACK TGNTRGRGR GTVPPPPPP GAPRTGDI.PP AVGTGPPPG  
51 AASKPPIIE ILQPSPTK ALREARKAP AIPDSRGVD TQIPSTPG  
101 RQOTMPTPO RTSTEVDRD VNAGTVGTFI VPRQITHSK KTAPMANGRI  
151 VVNDHLITV DPEQTSIST RATQEOFNAM YEGVEDYGV NDRQMGILLN  
201 GLAWCIENG TSPNMGWV MMDGBOVY PIKPLIDNAV PTFQIMTHF  
251 SDIAEYTER RNRIKAYMPR YGLQRMILDM SIARYAFDF ELHSNTPVRA  
301 REAHQOMKAA ALKNAQRLE GLDGNVSTQE EDTERHTTID VTNNINLLG  
351 MRGVQ

!!AA\_SEQUENCE 1.0  
P1:E71669 - preprotein translocase secy chain (secy) RP639 - Rickettsia  
prokazekii  
C:Species: Rickettsia provazekii  
C:Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 26-Aug-1999  
R:Accession: E71669  
R:Authors: S.G.E.; Zomorodipour, A.; Andersson, J.O.; Slicheritz-Ponten, T.;  
Almark, U.C.M.; Podowski, R.M.; Naeslund, A.K.; Eriksson, A.S.; Winkler, H.H.;  
Kurland, C.G.  
A:Title: The genome sequence of Rickettsia provazekii and the origin of  
mitochondria  
A:Reference number: A71630; MID:99039499  
A:Accession: E71669  
A>Status: Preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-433 <AND>  
A:Cross-references: GB:AJ235272; GB:AJ235269; MID:93861033; PID:e1342923;  
PID:93861179  
A:Experimental source: strain Madrid E  
C:Genetics:  
A:Gene: secy, RP639  
C:Superfamily: preprotein translocase secy

E71669 Length: 433 February 11, 2000 15:51 Type: P Check: 3620 ..

1 MGONFSKSS NDVNRILFT LFMILICRFQ SPIPIGIDS IALNSVAENK  
51 QFQILGMFM LSGSLGRMS IFALAIWPIY TASIIQIMS VAKPPLNKK  
101 KEGTGKRRKI NQLSRYLTVL LASFOAYGVA LLSMWTNT GPVILLAGF  
151 FRVTYITLV VGTILMWIG EQITRGIGN GTSIIIFGI ISGVPSAIIIS  
201 MEIENKRAL SPLAIWYCI GVVLLAIIT FFEKQKRLV VOYPRQVGN  
251 KIYGEATHM PLKNTSGVI PRIFASSILL PPTILASISN SNSDYSMLT  
301 YLHGKRPYV ILVYVLMF ESFFYIAIVF NSEETANNLR KYAYIPGKR  
351 PKKNSDVFY YILRLVIG GYLSVTCYI PELIMNKYVI SLSLGTSLF  
401 IVVNVLDITM TOTQTYFSS KTEGLMKKIK LKN

!!AA\_SEQUENCE 1.0  
P1:G69632 - transcription repressor glcr - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 26-Aug-1999  
R:Accession: G69632  
R:Authors: Ogatawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo,  
V.; Betteiro, M.G.; Bessieres, P.; Bolocin, A.; Borchert, S.; Boriss, R.;  
Bourstiel, L.; Brans, A.; Braun, M.; Brignelli, S.C.; Bron, S.; Brouillet, S.;  
Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani,  
J.O.; Conerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denicot, F.; Devine, K.M.;

Duesterhoeft, A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;  
Fahnestock, G.; Feriati, E.

Nature 380, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funai, S.; Gallizzi,  
G.; Galleron, N.; Gilm, S.Y.; Glaser, P.; Goffeau, A.; Goldthly, E.J.; Grandi,  
G.; Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.;  
Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.;  
Joris, B.; Karamata, D.; Kasahara, Y.; Kleerr-Bianchard, M.; Klein, C.;  
Kobayashi, Y.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kunita, K.;  
Lapidus, A.; Lardinois, S.  
A:Authors: Lamber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda,  
S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.;  
Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogikawa, A.; Oudega,  
B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Potwolik, S.; Prescott,  
A.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Ray, M.; Reynolds,  
S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato,  
T.; Scanlon, E.  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska,  
A.; Seror, S.J.; Serrot, P.; Shin, B.S.; Solido, B.; Sorokin, A.; Tacconi, E.;  
Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka,  
T.; Terpestra, P.; Tognoni, A.; Tosio, V.; Uchlyama, S.; Vandenberg, M.;  
Vannier, F.; Vaasarotti, A.; Viari, A.; Wanduit, R.; Wedler, E.; Wedler, H.;  
Weltenegeger, T.; Winters, P.; Wipet, A.; Yamamoto, H.; Yamane, K.; Yasunoto,  
K.; Yata, K.; Yoshida, K.  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danachin, A.  
A:Title: The complete genome sequence of the gram-positive bacterium Bacillus  
subtilis.  
A:Reference number: A69580; MID:96044033  
A:Accession: G69632  
A>Status: Preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-258 <KUN>  
A:Cross-references: GB:Z99122; GB:AL009126; MID:92636029; PID:e184536;  
PID:92636155  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: glcr  
C:Superfamily: regulatory protein glcr

G69632 Length: 258 February 11, 2000 15:51 Type: P Check: 4284 ..

1 MYOERLVAI LDFLKHNR I TEOITCTILO VSRDTRARDL VKLEGNALIT  
51 RTGGAILPT VHOKIOSYSG RKTIVSEKN KIGSLASLI HOGDRIYIDA  
101 STTVQACAKH LNAVDCTVIN NSINLADVLIS DKEIEIYLL GSKLEENHF  
151 IYSSVIEKL SSYHVDKALI GVGISEHGI TINHEEDGW KKKMQQAKO  
201 VIALADHSLK GSTSFQYAE LNEIDLITD RLPRQAFCD LDRNGVELLV  
251 TEODEGRD

!!AA\_SEQUENCE 1.0  
P1:F69260 - nitrate ABC transporter, permease protein (ntrB-1) homoIog -  
Archaeoglobus fulgidus  
C:Species: Archaeoglobus fulgidus  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 26-Aug-1999  
R:Accession: F69260  
R:Authors: Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum,  
K.A.; Dodson, R.J.; Gilm, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.;  
Klavage, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush,  
J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirtness, E.F.; Dougherty, B.A.;  
McKenney, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.;  
Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.;  
McDonald, L.  
Nature 390, 364-370, 1997  
A:Authors: Uddelback, T.; Cotton, M.D.; Spriggs, T.; Attlich, P.; Kaine, B.P.;  
Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;  
Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.  
A:Title: The complete genome sequence of the hyperthermophilic,  
sulfate-reducing archaeon Archaeoglobus fulgidus.  
A:Reference number: A69250; MID:98049343



Mon Feb 14 08:07:32 2000

plc.cat

Page 101

A:Accession: F69260  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-244 <KUN>  
A:Cross-references: GB:AE001100; GB:AE000782; NID:g2689423; PID:g2650562;  
TIGR:AF0086  
C:Superfamily: Synechococcus nitrate transport protein ntb  
F69260 Length: 244 February 11, 2000 15:51 Type: P Check: 6950 ..  
1 MQASKADYKG YLITLALL IMAAASYLQ NPALNPDPV LALLTKEL  
51 ITHAVSLIR VVYSLALAS VALPAGILSR ERYVDATISP IYLLIPIPH  
101 IYLLPIYLL FGIGDLSRYL LIATLEFQI AVTRDPAKO VSDYVYSIL  
151 SLGASKIDY RHVILPAMP KILPALRISI GMAIAVLEFA ESFAITSGLG  
201 YLIDMSERA DYTMYAAT TMAIGFALY VIVESAERNY CRWL  
11AA-SEQUENCE 1.0  
PI:A70039 - ABC transporter (ATP-binding protein) homolog yfsc - Bacillus  
subtilis  
C:Species: Bacillus subtilis  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 28-May-1999  
C:Accession: A70001  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, J.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, K.M.; J.T.; Conerton, I.F.; Cummings, N.J.; Daniel, R.A.; Danilov, F.; Devane, K.M.; Diesthorpe, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghm, S.Y.; Glaser, P.; Goffeau, A.; Goldlighty, E.J.; Grandt, G.; Gusspehl, G.; Guy, B.J.; Haga, K.; Halech, J.F.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karmali, D.; Kasahara, Y.; Kletter-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lablans, A.; Lardinois, S.  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandembol, M.; Vannier, F.; Vassaret, A.; Viari, A.; Wandut, R.; Wedler, E.; Wedler, H.; Weltzenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MUID:98044033  
A:Accession: A70001  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-253 <KUN>  
A:Cross-references: GB:499119; GB:AL009126; NID:g2635411; PID:el185911;  
PID:g2635522  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yfsc  
C:Superfamily: ATP-binding cassette homology  
C:Keywords: ATP; P-loop  
F:23-219/Domain: ATP-binding cassette homology <ABC>  
F:40-47/Region: nucleotide-binding motif A (P-loop)  
A70001 Length: 253 February 11, 2000 15:51 Type: P Check: 6987 ..  
1 MYLENKTR KSYGNKLNKO EYLGKIDHI EKGFYSIMQ ASGSGKTLIL  
51 NVISIDQVS HGTHINGD MTKAMKOLA EFRKHGFI FQYMLDLDTL  
101 TYKENILPL SITLKSKEA NKRFEVAKA LGEIENDKY PNEISGQKO  
151 RISAGRAFIH DSIIFADEP TGALDSKAS DLNKLQOLN QKRNATIMV  
201 THDPAASYC GRVIFIKDQ MTQLNKGO DROFFQDIM KQGVLGQVO  
251 HEH  
11AA-SEQUENCE 1.0  
PI:A70039 - ABC transporter (ATP-binding protein) homolog yfsc - Bacillus  
subtilis  
C:Species: Bacillus subtilis  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 28-May-1999  
C:Accession: A70039  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, J.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, K.M.; J.T.; Conerton, I.F.; Cummings, N.J.; Daniel, R.A.; Danilov, F.; Devane, K.M.; Diesthorpe, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghm, S.Y.; Glaser, P.; Goffeau, A.; Goldlighty, E.J.; Grandt, G.; Gusspehl, G.; Guy, B.J.; Haga, K.; Halech, J.F.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karmali, D.; Kasahara, Y.; Kletter-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lablans, A.; Lardinois, S.  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandembol, M.; Vannier, F.; Vassaret, A.; Viari, A.; Wandut, R.; Wedler, E.; Wedler, H.; Weltzenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MUID:98044033  
A:Accession: A70039  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-301 <KUN>  
A:Cross-references: GB:429121; GB:AL009126; NID:g2635827; PID:el186097;  
PID:g2635922  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yfsc  
C:Superfamily: ATP-binding cassette homology  
C:Keywords: ATP; P-loop  
F:21-206/Domain: ATP-binding cassette homology <ABC>  
F:38-45/Region: nucleotide-binding motif A (P-loop)  
A70039 Length: 301 February 11, 2000 15:51 Type: P Check: 1714 ..  
1 MKREAVTVSN YKHFORKTA VNNISFIEK GELIALIGN GAGRTVSM  
51 ILGLKBSG EKLNRNPD DQVRKIGV MGEVSVMPG LKVDLIELF

101 RSYNPPLSM KELVSLTALT KEDKTRAEK LSGGCKRRLS FALALANPE  
 151 LLIDEPVVG MDTSRRHRM QTINGLSOG KTIIFSRHYL QEADDAORI  
 201 LFTGGLVLA DGSPPMOIRSR IQKOSVSFTL HSESELERLS CPEVERVH  
 251 EHRTTIQTS NTDKVALIF QENIHARDIR IEQATIDEAF ROLADGNREA  
 301 M

!!AA\_SEQUENCE 1.0  
 P1:G69260 - nitrate ABC transporter, ATP-binding protein (ntrC-1) homolog -  
 Archaeoglobus fulgidus  
 C:Species: Archaeoglobus fulgidus  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 28-May-1999  
 C:Accession: G69260  
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kellavagge, A.R.; Graham, D.E.; Kyrleides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kinkness, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodex, A.; Zhou, L.; Overbeek, R.; Goeyne, J.D.; Weidman, J.F.; McDonald, L.  
 Nature 390, 364-370, 1997  
 A:Authors: Ullrich, T.; Cotton, M.D.; Springs, T.; Attleach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon Archaeoglobus fulgidus.  
 A:Reference number: A69250; MUID:98049343  
 A:Accession: G69260  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-243 <KLE>  
 A:Cross-references: GB:AE001100; GB:AE000782; NID:g2689423; PID:g2650561; TIGR:AF0087  
 C:Superfamily: ATP-binding cassette homology  
 C:Keywords: ATP; P-loop  
 F:17-205/Domain: ATP-binding cassette homology <ABC>  
 F:34-41/Region: nucleotide-binding motif A (P-loop)

G69260 Length: 243 February 11, 2000 15:51 Type: P Check: 9269 ..

1 MITVKGISKF FGLKALDGV SFEVDEGTC AITGPGCGK STILLIMAGL  
 51 LKPSGEVLV DRAVNSPK NAALIDDFG LFPKTYDYN VALGKIRGF  
 101 SRAERERTV ALLEKFLGK FEKSYKOLS GGMKORVVA RAIAPOLL  
 151 LMDPLSLD ALSREMONF LNLKERTKT TWLVTHSIE EAVFLGRKIV  
 201 VLTERPGRVK AAVDREAGD ESYREVEVF ERCKLRQTI RAT

!!AA\_SEQUENCE 1.0  
 P1:S56335 - hisP-like nucleotide binding protein phnK - Escherichia coli  
 N:Alternate names: phosphonates transport ATP-binding protein phnK  
 C:Species: Escherichia coli  
 C:Date: 28-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 28-May-1999  
 C:Accession: S56335; H65216; C35719  
 R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.  
 Nucleic Acids Res. 23, 2105-2119, 1995  
 A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.8 through 100 minutes.  
 A:Reference number: S56314; MUID:95334362  
 A:Accession: S56335  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-252 <BUR>  
 A:Cross-references: EMBL:U14003; NID:g1263172; PID:g5336941  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Colado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,

J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y. Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:97426617  
 A:Accession: H65216  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-252 <BLAT>  
 A:Cross-references: GB:AE000482; GB:U00096; NID:g2367349; PID:g1790535; UWGP:D4097  
 A:Experimental source: strain K-12, substrain MG1655  
 R:Chen, C.M.; Ye, Q.Z.; Zhu, Z.; Wanner, B.L.; Walsh, C.T.  
 J. Biol. Chem. 265, 4461-4471, 1990  
 A:Title: Molecular biology of carbon-phosphorus bond cleavage. Cloning and sequencing of the phn (psid) genes involved in alkylphosphate uptake and C-P lyase activity in Escherichia coli B.  
 A:Reference number: A35718; MUID:90170953  
 A:Accession: C35719  
 A:Molecule type: DNA  
 A:Residues: 1-46 / 1-48-252 <CHE>  
 A:Cross-references: GB:J05260; NID:g147192; PID:g147206  
 C:Genetics:  
 A:Gene: phnK  
 C:Superfamily: ATP-binding cassette homology  
 C:Keywords: ATP; P-loop  
 F:21-222/Domain: ATP-binding cassette homology <ABC>  
 F:38-45/Region: nucleotide-binding motif A (P-loop)

S56325 Length: 252 February 11, 2000 15:51 Type: P Check: 1039 ..

1 MNPLSLVNN LTHLYAPGK FSDVSFDLMP GEVLGIVES GSGKTTOLKS  
 51 ISARLTPQG EHYENRSY AMSEADRRL LTEMGVQHO HPLDGLRROV  
 101 SAGNIGERL MATGARHYGD IRATQKWE EYEIPANRID DLPTFGSGM  
 151 QQRQIARNL VTHPKLVEMD EPTGGLDVSV QARLDDLKG LVLELNLAIV  
 201 IVHDLGVAR LLAIDLIVK QGVYVESGLT DAVLDPPHNP YTOLVSSVL  
 251 QN

!!AA\_SEQUENCE 1.0  
 P1:S73545 - ATP-dependent RNA helicase dead - Mycoplasma pneumoniae (ATCC 29342) (SGC3)  
 N:Alternate names: hypothetical protein C12.orf450  
 C:Species: Mycoplasma pneumoniae  
 A:Variety: ATCC 29342  
 C:Date: 27-Feb-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Sep-1999  
 C:Accession: S73545  
 R:Himmelfreisch, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R.  
 Nucleic Acids Res. 24, 4420-4449, 1996  
 A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae.  
 A:Reference number: S73327; MUID:97105885  
 A:Accession: S73545  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-450 <HIM>  
 A:Cross-references: EMBL:AE000022; GB:U00089; NID:g1673882; PID:NAB95867.1; PID:g1673885  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996  
 C:Genetics:  
 A:Gene: dead  
 A:Genetic code: SGC3  
 C:Superfamily: unassigned DEAD/H box helicases; DEAD/H box helicase homology  
 C:Keywords: ATP; P-loop  
 F:47-342/Domain: DEAD/H box helicase homology <DEAD>  
 F:47-54/Region: nucleotide-binding motif A (P-loop)  
 F:150-155/Region: nucleotide-binding motif B  
 F:154-157/Region: DEAD motif

S73545 Length: 450 February 11, 2000 15:51 Type: P Check: 2944

1 MDSTFNLGV SPALIALTKD NNINOPTTIO QLAIPQLOH QNLIVHSPTG  
51 TGTAAVGGIP VIETLLKPKS KGTOTLVVA PYRELAEOIK TTEINFKAHT  
101 HMAVSLIGG IPIMQOLKOL ENOEIYVGT MGRVMDLLER GVIFEELHEH  
151 LIIDEDVDMC DRGFRRKFLD LLSRIEFEG IAVYSASVNE ETITAKOIT  
201 KNGIFLAPE LKONAPEDPN KLIDQFVCL FSNRKKQALY SLVQSTAKS  
251 IYPCDTRKL VDELCIFLRK NDVKTYPHIG DKAQFIRENN LKLANPTAP  
301 IYLVITDLIG RGIHVEGDM VVNSACVNF ETYLHMGRT GRNHNKSCSI  
351 TFCISHKQA FLKLEQVND KRISPLRPMR LRILPLKCTI QPKGKXLSLO  
401 SVOKIYVNR SNGTRRVPPL AGDLFSRMR QPRDMQKNK LHSDMOSNM

11AA\_SEQUENCE 1.0  
P1:566920 - Probable RNA helicase CAS/6 - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein O2774; protein YOR046c  
C:Species: Saccharomyces cerevisiae  
C>Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 20-Sep-1999  
C:Accession: S66920; S66929; D34848  
R:Landt, O.; Hiesel, R.; Unseid, M.  
Submitted to the Protein Sequence Database, July 1996  
A:Reference number: S66907  
A:Accession: S66920  
A:Molecule type: DNA  
A:Residues: 1-482 <LAN>  
A:Cross-references: EMBL:Z74954; NID:G1420174; PID:CAA9237.1; PID:E252333;  
PID:G1420175; MIPS:YOR046c  
A:Experimental source: strain S288C  
R:Bohn, C.; Bolestin-Fukuhara, M.; Daignan-Fornier, B.; Dang, D.V.; Valens, M.  
Submitted to the Protein Sequence Database, July 1996  
A:Reference number: S66929  
A:Accession: S66929  
A:Molecule type: DNA  
A:Residues: 1-246 <BOH>  
A:Cross-references: EMBL:Z74954; MIPS:YOR046c  
A:Experimental source: strain S288C  
R:Chang, T.H.; Arenas, J.; Abelson, J.  
Proc. Natl. Acad. Sci. U.S.A. 87:1571-1575, 1990  
A:Title: Identification of five putative yeast RNA helicase genes.  
A:Reference number: A34848; M01D:90180368  
A:Accession: D34848  
A:Molecule type: DNA  
A:Residues: 237-425 <CHA>  
C:Genetics:  
A:Gene: SGD:DBP5  
A:Cross-references: SGD:S0005572; MIPS:YOR046c  
A:Map position: 15R  
C:Superfamily: unassigned DEAD/H box helicases; DEAD/H box helicase homology  
C:Keywords: ATP; P-loop  
F:138-429/Domain: DEAD/H box helicase homology <DEAD>  
F:138-145/Region: nucleotide-binding motif A (P-loop)  
F:235-240/Region: nucleotide-binding motif B  
F:239-242/Region: DEAD motif

S66920 Length: 482 February 11, 2000 15:51 Type: P Check: 2958

1 MSDDRDPAD LLASLKIDNE KEDTSEVSTK ETVASQPEKT ADSIRPAEKL  
51 VKVEEKKTK QEDSNLISSE YEAVKVLADI QADPNSPLYS ASFDELIGA  
101 PELLGIGIAM KFOKPSKIOE RALPPLLHNP PRNNIAOSOS GTGKTAASL  
151 TMLTAVNPD ASPQALCLAP SRELARQILE VVQDMGKFK ITSQILVDS  
201 FKKKQINAO VIVGTPTVL DLMRRKMLQI QKIIIFVDE APMMDQOGL

251 GDQCIKRRF LPRDQVLE SATPADAVRQ YAKKIIVPANN TLEIQTENV  
301 VDAIKQLYMD CKNKADKEDV LTELGLMTI GSGIIFVATK KTANVLYGKL  
351 KSEGEVSIIL HGDLOQERD RLIDDFREGK SKVLTITNVL ARGIDIFTVS  
401 MVVNYDPLTL ANGOADPATY IHRIGRTGRF GRGVAISFV HDKSNFMLS  
451 AIQKRYGDIE MTRVPTDMD EVERIKVKVL KD

11AA\_SEQUENCE 1.0  
P1:A38900 - T-cell-specific transcription factor-1 - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 10-Mar-1994 #sequence\_revision 10-Mar-1994 #text\_change 08-Sep-1997  
C:Accession: A38900  
R:Oosterwegel, M.; van de Wetering, M.; Dooljes, D.; Klomp, L.; Winoto, A.; Georgopoulos, K.; Meljlink, F.; Clevers, H.  
J. Exp. Med. 173, 1133-1142, 1991  
A:Title: Cloning of murine TCF-1, a T cell-specific transcription factor interacting with functional motifs in the CD3-epsilon and T cell receptor alpha enhancers.  
A:Reference number: S17398; M01D:91217625  
A:Accession: A38900  
A:Molecule type: mRNA  
A:Residues: 1-272 <OOS>  
A:Experimental source: T-cell, cell line  
C:Superfamily: unassigned HMG box proteins; HMG box homology  
C:Keywords: transcription factor  
F:133-237/Domain: DNA binding #status predicted <DNA>  
F:154-229/Domain: HMG box homology <HMG1>

A38900 Length: 272 February 11, 2000 15:51 Type: P Check: 3880

1 MYKETVYSAF NLMPYPYPAS GAGQHPQOP PLHNKQOPQ HGVPLSLYL  
51 EHFSSPHPT APADISOKOG VHRPLQTPDL SGFYSLTGS MGQLPHTVSM  
101 FTHPSLMGS GVGHPAIP HPAIVPSSGQ QELQPYDRNL KTOAPEKAK  
151 EAKRPVIRKP LNAFMLYMEK MRKVIAECT LKESAINOI LGRRWHLASR  
201 EQQAYVELA RKEROLHML YPGMSARDNY GKRRRSREK HOESTTGKR  
251 NAFGTPEKA AAPAFELPMT VL

11AA\_SEQUENCE 1.0  
P1:JH0402 - T-cell-specific transcription factor-1D - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 10-Mar-1994 #sequence\_revision 10-Mar-1994 #text\_change 08-Sep-1997  
C:Accession: JH0402  
R:Oosterwegel, M.; van de Wetering, M.; Doeljes, D.; Klomp, L.; Winoto, A.; Georgopoulos, K.; Meljlink, F.; Clevers, H.  
J. Exp. Med. 173, 1133-1142, 1991  
A:Title: Cloning of murine TCF-1, a T cell-specific transcription factor interacting with functional motifs in the CD3-epsilon and T cell receptor alpha enhancers.  
A:Reference number: S17398; M01D:91217625  
A:Accession: JH0402  
A:Molecule type: mRNA  
A:Residues: 1-279 <OOS>  
A:Experimental source: T-cell, EL-4  
C:Superfamily: unassigned HMG box proteins; HMG box homology  
C:Keywords: transcription factor  
F:154-229/Domain: HMG box homology <HMG1>

JH0402 Length: 279 February 11, 2000 15:51 Type: P Check: 9457

1 MYKETVYSAF NLMPYPYPAS GAGQHPQOP PLHNKQOPQ HGVPLSLYL  
51 EHFSSPHPT APADISOKOG VHRPLQTPDL SGFYSLTGS MGQLPHTVSM  
101 FTHPSLMGS GVGHPAIP HPAIVPSSGQ QELQPYDRNL KTOAPEKAK

```

151 EAKRPVKRP LNAFLYKKE MRAKVIACET LKESAINOI LGRRWHLNR
201 EEOAKYIELA RKRRLHML YPGMSARDNY GKKRRSRK HOESTJNSL
251 HXSGKRAAF GTPEKAAP APFLPMVL

!!AA_SEQUENCE 1.0
PI:S50068 - nonhistone chromosomal protein HMG1-like protein - fruit fly
(Drosophila melanogaster)
N:Alternate names: dorsal switch protein
C:Species: Drosophila melanogaster
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 24-Sep-1995
C:Accession: S50068; S47596
R:Lehming, N.
A:Reference number: S50068
A:Reference number: S50068
A:Accession: S50068
A:Molecule type: mRNA
A:Residues: 1-393 <LEW>
A:Cross-references: EMBL:U13881; NID:9551089; PID:AA50238.1; PID:9551090
R:Lehming, N.; Thanos, D.; Brickman, J.M.; Ma, J.; Maniatis, T.; Ptashe, M.
A:Title: An HMG-like protein that can switch a transcriptional activator to a
repressor.
A:Reference number: S47596; MUID:94352416
A:Accession: S47596
A:Molecule type: mRNA
A:Residues: 178-393 <LEW>
A:Cross-references: EMBL:U13881
C:Genetics:
A:Gene: FLYBase:Dspl
A:Cross-references: FLYBase:Fbgn011764
C:Superfamily: unassigned HMG box proteins; HMG box homology
F:6-47/Region: glutamine-rich
F:131-161/Region: glutamine-rich
F:176-253/Domain: HMG box homology <HMG1>
F:268-342/Domain: HMG box homology <HMG2>

S50068 Length: 393 February 11, 2000 15:51 Type: P Check: 241 ..

1 MEHFHQIQCT IOHYQQQLAA QQQQQVQQQQ LQGHQVYQQ NOQQAHOSS
51 NTAGVGTQQ LFTYKASSE PNPAITMAQV VASNSAGTI GVDYLNMAQ
101 AAAAAVPGS QMWYSANQG QVDANTATL QHQQQQQQQQ QQQQQQHQQ
151 QQQQQQQQQ QWVNSAPM SRVADAKPR GRMTAYAYFV OTCREHKK
201 HPDEYIAFE FSRKCAERKK TMVDKEKRF HEKAEKDKOR YEAEQNTVP
251 PKALMDRKK KKKQIKDPNA PKRSLAPFW FCNDEKKNK ALNPEFGVD
301 IAKELGRKS DYDEYKQY ESMAERDKAR YEREKTEYT SGKIASAPS
351 MQASMQAQAO KALLAARAQ QOHQOLEQH DDDGDGDDDD ENQ

!!AA_SEQUENCE 1.0
PI:A38095 - T-cell-specific transcription factor 1 splice form A - human
N:Alternate names: transcription factor TCF-1A
C:Species: Homo sapiens (man)
C>Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 24-Sep-1995
C:Accession: A38095; S16645; S65005
R:van de Wetering, M.; Oosterwegel, M.; Holstege, F.; Dooyes, D.; Suijkerbuijk,
J.; Geurts van Kessel, A.; Clevers, H.
A:Title: The human T cell transcription factor-1 gene. Structure, localization,
and promoter characterization.
A:Reference number: A38095; MUID:92235082
A:Accession: A38095
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-269 <VNA>
A:Cross-references: GB:X63901

A:Note: authors translated the codon GAT for residue 253 as Ser
R:van de Wetering, M.; Oosterwegel, M.; Dooyes, D.; Clevers, H.
EMO v. 10, 123-132, 1991
A:Title: Identification and cloning of TCF-1, a T lymphocyte-specific
transcription factor containing a sequence-specific HMG box.
A:Reference number: S13449; MUID:91114695
A:Accession: S16645
A:Molecule type: mRNA
A:Residues: 1-269 <WEN>
A:Cross-references: EMBL:X59869; NID:936785; PID:CAA42526.1; PID:936786
R:Mayer, K.; Wolff, E.; Clevers, H.; Balhansen, W.G.
Biochim. Biophys. Acta 1263, 169-172, 1995
A:Title: The human high mobility group (HMG)-box transcription factor TCF-1:
novel isoforms due to alternative splicing and usage of a new exon IXA.
A:Reference number: S61796; MUID:95367594
A:Accession: S65005
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 244-269 <MAX>
C:Genetics:
A:Introns: 32/3; 68/1; 97/2; 137/2; 191/3; 227/3; 244/1
C:Superfamily: unassigned HMG box proteins; HMG box homology
C:Keywords: alternative splicing; DNA binding; transcription factor
F:151-226/Domain: HMG box homology <HMG1>

A38095 Length: 269 February 11, 2000 15:51 Type: P Check: 5007 ..

1 MKKRYIAF NLMMHYPPS GAGQHPQPOP PLRKANOPH GVPOLSYEH
51 FNSRPAPAD ADISQKQVR PLQTPDLSGF YSLTSGMQ LPHYVSWTH
101 PSLMGSSVP GHPAIPHPA IVPSPGQEL QPFDRLTKT AESKAEKAK
151 KPTTKPLNA EMLYKEMRA KVIAECTLKE SAAINQILGR RWHALSREO
201 AKYELARKE ROLHMLYPG WSRDNYGK KRSREKHQE STTETNWPBE
251 LKDGNGQEST SMSSSSSPA

!!AA_SEQUENCE 1.0
PI:A38095 - T-cell-specific transcription factor 1 splice form B - human
N:Alternate names: transcription factor TCF-1B
C:Species: Homo sapiens (man)
C>Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 24-Sep-1995
C:Accession: A38095; S78559; S65006
R:van de Wetering, M.; Oosterwegel, M.; Holstege, F.; Dooyes, D.; Suijkerbuijk,
J.; Geurts van Kessel, A.; Clevers, H.
A:Title: The human T cell transcription factor-1 gene. Structure, localization,
and promoter characterization
A:Reference number: A38095; MUID:92235082
A:Accession: A38095
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-269 <VNA>
A:Cross-references: GB:X63901
R:van de Wetering, M.; Oosterwegel, M.; Dooyes, D.; Clevers, H.
EMO v. 10, 123-132, 1991
A:Title: Identification and cloning of TCF-1, a T lymphocyte-specific
transcription factor containing a sequence-specific HMG box.
A:Reference number: S13449; MUID:91114695
A:Accession: S78559
A:Molecule type: mRNA
A:Residues: 1-269 <VNA>
A:Cross-references: EMBL:X59870; NID:936787; PID:CAA42527.1; PID:936788
R:Mayer, K.; Wolff, E.; Clevers, H.; Balhansen, W.G.
Biochim. Biophys. Acta 1263, 169-172, 1995
A:Title: The human high mobility group (HMG)-box transcription factor TCF-1:
novel isoforms due to alternative splicing and usage of a new exon IXA.
A:Reference number: S61796; MUID:95367594
A:Accession: S65006
A:Status: not compared with conceptual translation
A:Molecule type: mRNA

```

A:Residues: 244-269 <MAY>  
A>Note: only a part of the coding sequence is given  
C:Genetics:  
A:Insertions: 32/3; 68/1; 97/2; 137/2; 191/3; 227/3; 243/3  
N:Alternate names: transcription factor TCF-1C  
C:Species: Homo sapiens (man)  
C>Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 24-Sep-1999  
C:Accession: D38095; S61877; S61800  
R:van de Wetering, M.; Oosterwegel, M.; Holstege, F.; Dooyes, D.; Suljkerbuijk, R.; Geurts van Kessel, A.; Clevers, H.  
J. Biol. Chem. 267, 8530-8536, 1992  
A:Title: The human T cell transcription factor-1 gene. Structure, localization, and promoter characterization  
A:Reference number: A38095; MUID:92235082  
A:Accession: D38095  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-250 <MAY>  
A:Cross-references: GB:X63901  
R:Mayer, K.; Wolff, E.; Clevers, H.; Ballhausen, W.G.  
submitted to the EMBL Data Library, January 1995  
A:Description: The human high mobility group (HMG)-box transcription factor TCF-1: novel isoforms due to alternative splicing and usage of a new exon IXA.  
A:Reference number: S61877  
A:Accession: S61877  
A:Molecule type: mRNA  
A:Residues: 1-250 <MAY>  
A:Cross-references: EMBL:247361; NID:9619881; PIDN:CAA87439.1; PID:9619882  
R:Mayer, K.; Wolff, E.; Clevers, H.; Ballhausen, W.G.  
Biochim. Biophys. Acta 1263, 169-172, 1995  
A:Title: The human high mobility group (HMG)-box transcription factor TCF-1: novel isoforms due to alternative splicing and usage of a new exon IXA.  
A:Reference number: S61796; MUID:95367594  
A:Accession: S61800  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 244-250 <MAW>  
A:Cross-references: EMBL:247361  
C:Genetics:  
A:Insertions: 32/3; 68/1; 97/2; 137/2; 191/3; 227/3; 243/3  
N:Alternate names: HMG box proteins; HMG box homology  
C:Keywords: alternative splicing; DNA binding; transcription factor  
F:151-226/Domain: HMG box homology <HMG1>  
D38095 Length: 269 February 11, 2000 15:51 Type: P Check: 3201 ..

1 MYKETVSAF NLNHYPPS GAGHPPOF PLHKANOPH GVPOLSLYH  
51 FNSPHPTAP ADISOKOVH PLQTPDLSCF YSLTSGSMQ LPHTVSWFH  
101 PSLMGSGVP GHPAIPHPA IVPSSGKOL QPDRNLKTO AESKAKEAK  
151 KPTIKKPLNA FMYLKEKRA KVIAECTIKE SAAINOLIGR RWHALSREDO  
201 AKYELARKE ROLHMOLYPG WSARDNYGKK KRSREKHOE STTGGRNMF  
251 GTYEKAAP APFLMTVL

!!AA\_SEQUENCE 1.0  
P1:D38095 - T-cell-specific transcription factor 1 splice form D - human  
N:Alternate names: transcription factor TCF-1D  
C:Species: Homo sapiens (man)  
C>Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 24-Sep-1999  
C:Accession: D38095; S61877; S61800  
R:van de Wetering, M.; Oosterwegel, M.; Holstege, F.; Dooyes, D.; Suljkerbuijk, R.; Geurts van Kessel, A.; Clevers, H.  
J. Biol. Chem. 267, 8530-8536, 1992  
A:Title: The human T cell transcription factor-1 gene. Structure, localization, and promoter characterization  
A:Reference number: A38095; MUID:92235082  
A:Accession: D38095  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-250 <MAY>  
A:Cross-references: GB:X63901  
R:van de Wetering, M.; Oosterwegel, M.; Dooyes, D.; Clevers, H.  
EMBO J. 10, 123-132, 1991  
A:Title: Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box.  
A:Reference number: S13449; MUID:91114695  
A:Accession: S13449  
A:Molecule type: mRNA  
A:Residues: 1-268 <WET>  
A:Cross-references: EMBL:X59871; NID:936789; PIDN:CAA42528.1; PID:936790  
A>Note: the authors describe an additional C-terminal exon of splice form C, the which is not contained in the respective nucleic acid sequence submitted to the EMBL data library and not mentioned in references A38095 and S61796  
R:Mayer, K.; Wolff, E.; Clevers, H.; Ballhausen, W.G.  
Biochim. Biophys. Acta 1263, 169-172, 1995  
A:Title: The human high mobility group (HMG)-box transcription factor TCF-1: novel isoforms due to alternative splicing and usage of a new exon IXA.  
A:Reference number: S61796; MUID:95367594  
A:Accession: S65007  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 244-268 <MAY>  
A>Note: only a part of the coding sequence is given  
C:Genetics:  
A:Insertions: 32/3; 68/1; 97/2; 137/2; 191/3; 227/3 244/1  
C:Superfamily: unassigned HMG box proteins; HMG box homology  
C:Keywords: alternative splicing; DNA binding; transcription factor  
F:151-226/Domain: HMG box homology <HMG1>  
C38095 Length: 268 February 11, 2000 15:51 Type: P Check: 409 ..

201 AKYELARKE ROLHMOLYPG WSARDNYGKK KRSREKHOE STTDSNLHYS  
!!AA\_SEQUENCE 1.0  
P1:C38095 - T-cell-specific transcription factor 1 splice form C - human  
N:Alternate names: transcription factor TCF-1C  
C:Species: Homo sapiens (man)  
C>Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 24-Sep-1999  
C:Accession: C38095; S13449; S65007  
R:van de Wetering, M.; Oosterwegel, M.; Holstege, F.; Dooyes, D.; Suljkerbuijk, R.; Geurts van Kessel, A.; Clevers, H.  
J. Biol. Chem. 267, 8530-8536, 1992  
A:Title: The human T cell transcription factor-1 gene. Structure, localization, and promoter characterization  
A:Reference number: A38095; MUID:92235082  
A:Accession: C38095  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-268 <VAN>  
A:Cross-references: GB:X63901  
R:van de Wetering, M.; Oosterwegel, M.; Dooyes, D.; Clevers, H.  
EMBO J. 10, 123-132, 1991  
A:Title: Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box.  
A:Reference number: S13449; MUID:91114695  
A:Accession: S13449  
A:Molecule type: mRNA  
A:Residues: 1-268 <WET>  
A:Cross-references: EMBL:X59871; NID:936789; PIDN:CAA42528.1; PID:936790  
A>Note: the authors describe an additional C-terminal exon of splice form C, the which is not contained in the respective nucleic acid sequence submitted to the EMBL data library and not mentioned in references A38095 and S61796  
R:Mayer, K.; Wolff, E.; Clevers, H.; Ballhausen, W.G.  
Biochim. Biophys. Acta 1263, 169-172, 1995  
A:Title: The human high mobility group (HMG)-box transcription factor TCF-1: novel isoforms due to alternative splicing and usage of a new exon IXA.  
A:Reference number: S61796; MUID:95367594  
A:Accession: S65007  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 244-268 <MAY>  
A>Note: only a part of the coding sequence is given  
C:Genetics:  
A:Insertions: 32/3; 68/1; 97/2; 137/2; 191/3; 227/3 244/1  
C:Superfamily: unassigned HMG box proteins; HMG box homology  
C:Keywords: alternative splicing; DNA binding; transcription factor  
F:151-226/Domain: HMG box homology <HMG1>  
C38095 Length: 268 February 11, 2000 15:51 Type: P Check: 409 ..

1 MYKETVSAF NLNHYPPS GAGHPPOF PLHKANOPH GVPOLSLYH  
51 FNSPHPTAP ADISOKOVH PLQTPDLSCF YSLTSGSMQ LPHTVSWFH  
101 PSLMGSGVP GHPAIPHPA IVPSSGKOL QPDRNLKTO AESKAKEAK  
151 KPTIKKPLNA FMYLKEKRA KVIAECTIKE SAAINOLIGR RWHALSREDO  
201 AKYELARKE ROLHMOLYPG WSARDNYGKK KRSREKHOE STDPGSPFK  
251 CRAREGLMOQ TDMCGPCR

!!AA\_SEQUENCE 1.0  
P1:JC6179 - dorsal switch protein 1 - fruit fly (Drosophila melanogaster)  
N:Alternate names: Drosophila melanogaster  
C:Species: Drosophila melanogaster  
C>Date: 11-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 24-Sep-1999  
C:Accession: JC6179  
R:Campanile, J.; Decoville, M.; Leng, M.; Locker, D.  
Gene 184, 285-290, 1997  
A:Title: The Drosophila DSL gene encoding an HMG 1-like protein: genomic organization, evolutionary conservation and expression.  
A:Reference number: JC6179; MUID:97103675  
A:Accession: JC6179

A:Molecule type: mRNA  
 A:Residues: 1-393 <CAN>  
 A:Cross-references: EMBL:X89811; NID:G1150374; PIDN:CAA61938.1; PID:el90037;  
 PID:G1150375  
 C:Superfamily: unassigned HMG box proteins; HMG box homology  
 F:116-253/Domain: HMG box homology <HMG1>  
 F:268-342/Domain: HMG box homology <HMG2>

J06179 Length: 393 February 11, 2000 15:51 Type: P Check: 83

1 MEHFQIOT IOHYOOLAA OOOOQVOOQ LOOHQVYVQ NOQOANONS  
 51 NTAGVGTOQ LFTYKASFF PNPATMAOV VATSNAAGTI GYDRLNMAQ  
 101 AAAAAVPGS QMWYSAANG QVDANTAAQL QHQOQOQOQ QOQOQOQOQ  
 151 QOOHQOQOQ QVYNSASPM SVKADAKPR GRMTAVYEV QTCREHKKK  
 201 HPDETIVFAE FSRKCAERWK TWVDEKKRF HEMAERKOR YEAMONTYP  
 251 PKGAVGKRG KRQIKDPNA PKRSIAEFW FCNDERKVK ALNPEFGVD  
 301 IAKELGRKMS DVDEYKOKY ESMERDKAR YEREMTEYKT SKRIAMSAPS  
 351 MQAMQMOAQ KALLALAAAO QOHQOLEQH DDDGDGDDDD ENQ

!!AA\_SEQUENCE 1.0  
 P1:S61796 - T-cell-specific transcription factor 1 splice form E - human  
 N:Alternate names: transcription factor TCF-1E  
 C:Species: Homo sapiens (man)  
 C>Date: 28-Oct-1996 #sequence\_revision 24-Jul-1998 #text\_change 24-Sep-1999  
 C:Accession: S61796; S61878  
 R:Mayer, R.; Wolff, E.; Clevers, H.; Ballhausen, W.G.  
 Biochem. Biophys. Acta 1263, 169-172, 1995  
 A:Title: The human high mobility group (HMG)-box transcription factor TCF-1:  
 novel isoforms due to alternative splicing and usage of a new exon IXA.  
 A:Reference number: S61796; MUID:95367594  
 A:Accession: S61796  
 A>Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 244-366 <MAY>  
 A:Note: only a part of the coding sequence is given  
 R:Mayer, R.; Wolff, E.; Clevers, H.; Ballhausen, W.G.  
 submitted to the EMBL Data Library, January 1995  
 A:Description: The human high mobility group (HMG)-box transcription factor  
 TCF-1: novel isoforms due to alternative splicing and usage of a new exon IXA.  
 A:Reference number: S61877  
 A:Accession: S61878  
 A:Molecule type: mRNA  
 A:Residues: 1-324, 'RHLPQVPLSASQPGPHRPPAACRAHRSNNRLNDRWPSRRHRTGRLQEP',  
 <MAY>  
 A:Cross-references: EMBL:Z47362; NID:G619883; PIDN:CAA87440.1; PID:G619884  
 A:Note: the difference in residues C-terminal of 324 is due to a frameshift  
 error  
 C:Superfamily: unassigned HMG box proteins; HMG box homology  
 C:Keywords: alternative splicing; DNA binding; transcription factor  
 F:151-226/Domain: HMG box homology <HMG1>

S61796 Length: 366 February 11, 2000 15:51 Type: P Check: 5713

1 MKETVYAF NLMMHYPPS GAGOHPOPO PLHKAPOPH GVPQLSYEH  
 51 FNSPHPTAP ADISQOVNR PLQTPDLGQ YSLTSGSGNQ LPHVSMETH  
 101 PSLMGSGVP GHPAIPHPA IYPPSGKQEL QPDRNLKTO ABEKAKKAK  
 151 KPTIKKPLNA EMLYKEMRA KYIACETLKE SAALNQLGR RMHALSREDO  
 201 AKYELARKE ROLHMLYGP WSARDNYKK KRRREKHQE SITDPSPPK  
 251 CRAFGINQO TDMCGPCRRK KKCIIYLGE GRCSPPSPD DSALGCPGSP  
 301 APOSPSYHL LPRFTEILT SPAEAPISP GLSTALSLEPT PGPPAPNST

### 351 LOSTOVOOE SOROVA

!!AA\_SEQUENCE 1.0  
 P1:J07677 homeobox protein HEX - human  
 C:Species: Homo sapiens (man)  
 C>Date: 30-Sep-1993 #sequence\_revision 20-Aug-1994 #text\_change 24-Sep-1999  
 C:Accession: J07677; S26799; S78048; S30231  
 R:Hiromas, R.; Radich, J.; Collins, S.  
 Biochem. Biophys. Res. Commun. 195, 976-983, 1993  
 A:Title: PCR cloning of an orphan homeobox gene (PRH) preferentially expressed  
 in myeloid and liver cells.  
 A:Reference number: J07677; MUID:93384629  
 A:Accession: J07677  
 A:Molecule type: DNA  
 A:Residues: 1-270 <ACC>  
 A:Cross-references: GB:L16499; NID:G292404; PIDN:AA02988.1; PID:G292405  
 R:Crompton, M.R.; Bartlett, T.J.; MacGregor, A.D.; Manfioletti, G.; Burattini,  
 E.; Giannotti, V.; Goodwin, G.H.  
 Nucleic Acids Res. 20, 5661-5667, 1992  
 A:Title: Identification of a novel vertebrate homeobox gene expressed in  
 haematopoietic cells.  
 A:Reference number: S26799; MUID:93087175  
 A:Accession: S26799  
 A:Molecule type: mRNA  
 A:Residues: 1-114, 'L', '116-270 <CRO>  
 A:Cross-references: EMBL:X67235; NID:G32547; PIDN:CAA47661.1; PID:G32548  
 R:Bedford, F.K.  
 submitted to the EMBL Data Library, February 1993  
 A:Reference number: S78048  
 A:Accession: S78048  
 A:Molecule type: DNA  
 A:Residues: 122-270 <BED>  
 A:Cross-references: EMBL:Z21533; NID:G32068; PIDN:CAA79730.1; PID:G32069  
 R:Bedford, F.K.; Asmuth, A.; Enver, T.; Wiedemann, L.M.  
 Nucleic Acids Res. 21, 1245-1249, 1993  
 A:Title: HEX: a novel homeobox gene expressed during haematopoiesis and  
 conserved between mouse and human.  
 A:Reference number: S30230; MUID:93219088  
 A:Accession: S30231  
 A:Molecule type: DNA  
 A:Residues: 136-195 <BEW>  
 A:Cross-references: EMBL:Z21533  
 C:Comment: This protein is an important regulator of normal hematopoiesis.  
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
 F:138-194/Domain: homeobox homology <HOX>

J07677 Length: 270 February 11, 2000 15:51 Type: P Check: 3795

1 MGYPHGPAA GAVGVPLXAP TPLQDPAHPT PFYEDILGR GRAPAPDAP  
 51 LPSNSSFST LVSPYRTPVY EPLPHPAFS HHSAAALAA YGPGGFGPL  
 101 YFPRTVNDY THALVRDPL GKPLMSPL QRLPHKRGQ QVRFSDQTI  
 151 ELKKFETOK YLSPERKRL AKMLQSERQ VKTWFONRRA KWRRLKQENP  
 201 QSNKKELES LDSSCDOROD LPSEQNKGS LDSSQCSPP ASQEDLESEI  
 251 SEDSDQEVDI EGDKSYFNAG

!!AA\_SEQUENCE 1.0  
 P1:I46089 - thyroid transcription factor-1 - dog  
 C:Species: Canis lupus familiaris (dog)  
 C>Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 24-Sep-1999  
 C:Accession: I46089; S42104  
 R:Van Renterghem, P.H.G.; Dremier, S.; Vassari, G.; Christophe, J.  
 Mol. Cell. Endocrinol. 112, 83-93, 1995  
 A:Title: Study of TTF1 gene expression in dog thyrocytes in primary culture.  
 A:Reference number: I46089; MUID:96034516

A:Accession: 146089  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-371 <VAN>  
 A:Cross-references: EMBL:X77910; NID:9457488; PIDN:CAA54868.1; PID:9457489  
 C:Genetics:  
 A:Gene: TTF-1  
 C:Superfamily: unassigned homeobox proteins; homeobox homology  
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
 F:162-218/Domain: homeobox homology <HOX>  
 146089 Length: 371 February 11, 2000 15:51 Type: P Check: 4988 ..

1 MSMSRKHPTP FSVSDILSPL EESYKKGME GGLGAPLAA YRQGAAPPA  
 51 AAMQCHAVGH HGAVTAAYHM TAAGVPOLSH SAVGVCNGN LGNMSELPY  
 101 ODTMNSASG PGWYGANDP REPALSRFNG PASGMNSGM GGLGSLGDVS  
 151 KNAAPLPSPAP RRRRVLFSSQ AOVYELERF KQKYLAP REHLASMIHL  
 201 TPTQVIMFQ NHRKMKRQA KDKAAQOOLQ QDSGGGGGGG GAGCPOOQQA  
 251 QOOSPRRVA VPIVYKDKPC QAGAPAPGAG SLQGHAAQQA QOQAAQAQA  
 301 AAALSVGSG PGLAHGPHQ PGSAGQSPDL AHHAASPAAL QGVVSLSHL  
 351 NSSGSDYGM SCSTLLYGR W

11AA\_SEQUENCE 1.0  
 P1:530230 - homeotic protein Hex - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 24-Sep-1999  
 C:Accession: S30230  
 R:Bedford, F.K.; Ashworth, A.; Ewyer, T.; Wiedemann, L.M.  
 Nucleic Acids Res. 21, 1245-1249, 1993  
 A:Title: HEX: a novel homeobox gene expressed during haematopoiesis and conserved between mouse and human  
 A:Reference number: S30230; M0ID:95219088  
 A:Accession: S30230  
 A:Molecule type: DNA  
 A:Residues: 1-271 <BED>  
 A:Cross-references: EMBL:Z21524; NID:9288500; PIDN:CAA9729.1; PID:9288501  
 C:Superfamily: unassigned homeobox proteins; homeobox homology  
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
 F:139-195/Domain: homeobox homology <HOX>  
 S30230 Length: 271 February 11, 2000 15:51 Type: P Check: 8616 ..

1 MOFPHGPAA APAGVPLXA PTPILQPAHP TPYIDILG RGPAAFPPTP  
 51 TLPSNNSFT SLVSSYRTPV YEPTRVHPAF SHHPAALAA AYGPSGEGCP  
 101 LYPFRITVND YTHALRHDP LCKPLWSPF LQRPILNRKG GQVRESNDOT  
 151 VLEKKFETQ KYLSPPERKR LAKMLQLSER QVKTWFQNR AKWRILKQEN  
 201 POSNKKDALD SLDTSCGOQ DLPSRONKGA SLDRSQCSPP PASQEDPDSE  
 251 IESDSQEVQ IESDKGFNA G

11AA\_SEQUENCE 1.0  
 P1:53724 - thyroid-specific enhancer-binding protein (T/EBP) - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 23-Aug-1995 #sequence\_revision 19-Oct-1995 #text\_change 24-Sep-1999  
 C:Accession: S53724  
 R:Oguchi, H.; Pan, Y.T.; Kimura, S.  
 Biochim. Biophys. Acta 1261, 304-306, 1995  
 A:Title: The complete nucleotide sequence of the mouse thyroid-specific enhancer-binding protein (T/EBP) gene: extensive identity of the deduced amino acid sequence with the human protein.  
 A:Reference number: S53724; M0ID:95226463  
 A:Accession: S53724

A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-372 <OGU>  
 A:Cross-references: GB:U19755; NID:9885887; PIDN:AAA86100.1; PID:9885888  
 C:Superfamily: unassigned homeobox proteins; homeobox homology  
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
 F:162-218/Domain: homeobox homology <HOX>  
 S53724 Length: 372 February 11, 2000 15:51 Type: P Check: 7622 ..

1 MSMSRKHPTP FSVSDILSPL EESYKKGME GGLGAPLAA YRQGAAPPA  
 51 AAMQCHAVGH HGAVTAAYHM TAAGVPOLSH SAVGVCNGN LGNMSELPY  
 101 ODTMNSASG PGWYGANDP REPALSRFNG PASGMNSGM GGLGSLGDVS  
 151 KNAAPLPSPAP RRRRVLFSSQ AOVYELERF KQKYLAP REHLASMIHL  
 201 TPTQVIMFQ NHRKMKRQA KDKAAQOOLQ QDSGGGGGGG GAGCPOOQQA  
 251 QOOSPRRVA VPIVYKDKPC QAGAPAPGAG SLQGHAAQQA QOQAAQAQA  
 301 AAALSVGSG PGLAHGPHQ PGSAGQSPDL AHHAASPAAL QGVVSLSHL  
 351 NSSGSDYGM SCSTLLYGR TW

11AA\_SEQUENCE 1.0  
 P1:56451 - thyroid-specific enhancer-binding protein (T/EBP) - human  
 N:Alternate names: Drosophila NK-2 homolog A; thyroid nuclear factor; thyroid transcription factor 1  
 C:Species: Homo sapiens (man)  
 C:Date: 27-Oct-1995 #sequence\_revision 27-Oct-1995 #text\_change 24-Sep-1999  
 C:Accession: A56451; S53723; S53725; G02041  
 R:Reda, R.; Clark, J.C.; Shaw-White, J.R.; Stahlman, M.T.; Boulet, C.J.; Whitsett, J.A. 270, 8108-8114, 1995  
 J. Biol. Chem.  
 A:Title: Gene structure and expression of human thyroid transcription factor-1 in respiratory epithelial cells  
 A:Reference number: A56451; M0ID:95229626  
 A:Accession: A56451  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-371 <TKE>  
 A:Cross-references: GB:U19816; NID:9767832; PIDN:ANC50125.1; PID:9767833  
 R:Salardi, A.; Tassi, V.; de Filippis, V.; Civitareale, D.  
 Biochim. Biophys. Acta 1261, 307-310, 1995  
 A:Title: Cloning and sequence analysis of human thyroid transcription factor 1.  
 A:Reference number: S53723; M0ID:95226464  
 A:Accession: S53723  
 A:Status: preliminary  
 A:Molecule type: RNA  
 A:Residues: 1-371 <SAI>  
 A:Cross-references: EMBL:X82850; NID:9695582; PIDN:CAA58053.1; PID:9695583  
 R:Oguchi, H.; Pan, Y.T.; Kimura, S.  
 Biochim. Biophys. Acta 1261, 304-306, 1995  
 A:Title: The complete nucleotide sequence of the mouse thyroid-specific enhancer-binding protein (T/EBP) gene: extensive identity of the deduced amino acid sequence with the human protein.  
 A:Reference number: S53724; M0ID:95226463  
 A:Accession: S53725  
 A:Status: preliminary  
 A:Molecule type: RNA  
 A:Residues: 1-241, 'G', 242-371 <OGU>  
 R:Hamdan, H.  
 Submitted to the EMBL Data Library, August 1995  
 A:Reference number: G09102  
 A:Accession: G02041  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-225, 'HE', 228-371 <HAM>  
 A:Cross-references: EMBL:U03749; NID:91072047; PID:91072048  
 C:Genetics:  
 A:Gene: GDB:TTF1; NKX2A; TTF-1

A:Cross-references: GDB:132588; OMIM:600635  
A:Map position: 14q13-14q13  
A:Introns: 125/1  
C:Superfamily: unassigned homeobox proteins; homeobox homology  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:162-218/Domain: homeobox homology <HOX>  
A56451 Length: 371 February 11, 2000 15:51 Type: P Check: 5439 ..

1 MMSMKHTTP FSVSDILSPL EESYKRYGME GGLGAPLAA YRGOAAPP  
51 AAOQOAHVGH HGAVTAAAHM TAAGVPOLSH SAVGVCNCGN LGNNSELPY  
101 QDTMRNSASG PGWYGANPDP RFPAISRFG PASGMMNSGM GGLSLGDS  
151 KNNAPLPSPAR RKRRLVFSQ AOVELERRF KOKKYLSPAE REHLASMTL  
201 TPPOVKATWQ NHRKMKRQA KDKAAQOQLQ QDSGGGGGGG GTGCPQOQQA  
251 OQGSRRRAV PVLYKDGKPC QAGAPAPGAA SIQGHQOQA QHQAQAQQA  
301 AAASVGSQG AGLAGHPGHQ PGSAGQSPDL AHHAASPAAL QGVVSLSHL  
351 NSSGSDYGTW SCSTLLYGR T W

!!AA\_SEQUENCE 1.0  
P1:SI2002 - thyroid nuclear factor 1 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 13-Jan-1995 #sequence\_revision 13-Mar-1997 #text\_change 24-Sep-1999  
C:Accession: SI2002; PC2252  
R:Gazzzi, S.; Price, M.; de Felice, M.; Damante, G.; Mattel, M.G.; di Lauro, R.  
E:EMBL J. 9, 3631-3639, 1990  
A:Title: Thyroid nuclear factor 1 (TF-1) contains a homeodomain and displays a novel DNA binding specificity.  
A:Reference number: SI2002; MUID:91006063  
A:Accession: SI2002  
A:Molecule type: mRNA  
A:Residues: 1-372 <CU2>  
A:Cross-references: EMBL:X53658; NID:957422; PIDN:CAA37851.1; PID:957423  
A:Note: Met-1 was determined by sequencing DNA  
R:Endo, T.; Ohta, K.; Saito, T.; Harauchi, K.; Nakazato, M.; Kogai, T.; Onaya, T.  
Biochem. Biophys. Res. Commun. 204, 1358-1363, 1994  
A:Title: Structure of the rat thyroid transcription.  
A:Reference number: PC2252; MUID:95071477  
A:Accession: PC2252  
A:Molecule type: DNA  
A:Residues: 1-35 <EMD>  
A:Cross-references: DDBJ:D38035  
A:Experimental source: Liver  
C:Comment: This protein activates the genes of thyroid specific protein, thyroglobulin, thyroid peroxidase and thyrotropin receptor.  
C:Genetics: TTF-1  
C:Superfamily: unassigned homeobox proteins; homeobox homology  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:102-218/Domain: homeobox homology <HOX>  
SI2002 Length: 372 February 11, 2000 15:51 Type: P Check: 6608 ..

1 MMSMKHTTP FSVSDILSPL EESYKRYGME GGLGAPLAA YRGOAAPP  
51 AAOQOAHVGH HGAVTAAAHM TAAGVPOLSH SAVGVCNCGN LGNNSELPY  
101 QDTMRNSASG PGWYGANPDP RFPAISRFG PASGMMNSGM GGLSLGDS  
151 KNNAPLPSPAR RKRRLVFSQ AOVELERRF KOKKYLSPAE REHLASMTL  
201 TPPOVKATWQ NHRKMKRQA KDKAAQOQLQ QDSGGGGGGG GTGCPQOQQA  
251 OQGSRRRAV PVLYKDGKPC QAGAPAPGAA SIQGHQOQA QHQAQAQQA  
301 AAASVGSQG AGLAGHPGHQ PGSAGQSPDL AHHAASPAAL QGVVSLSHL  
351 NSSGSDYGTW SCSTLLYGR T W

351 LNSSGSDYGA MSCSTLLYGR TW

!!AA\_SEQUENCE 1.0  
P1:I59234 - octamer binding transcription factor 1 - human  
C:Species: Homo sapiens (man)  
C>Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 24-Sep-1999  
C:Accession: I59234  
R:Bhargava, A.K.; Li, Z.; Weissman, S.M.  
Proc. Natl. Acad. Sci. U.S.A. 90, 10260-10264, 1993  
A:Title: Differential expression of four members of the POU family of proteins in activated and phorbol 12-myristate 13-acetate-treated Jurkat T cells.  
A:Reference number: I59234; MUID:94052142  
A:Accession: I59234  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-420 <RES>  
A:Cross-references: GB:I20433; NID:9418015; PIDN:AAA65605.1; PID:9418016  
C:Superfamily: unassigned homeobox proteins; homeobox homology; POU domain homology  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:268-338/Domain: POU domain homology <POU>  
F:357-413/Domain: homeobox homology <HOX>  
I59234 Length: 420 February 11, 2000 15:51 Type: P Check: 9010 ..

1 MMSMKOPH FMAHTLPEH KYPSSLSSSE AIRRACLPPT PLOSNLFASTL  
51 DETLLARAA LAADVAVSQ GKSHPEKPPA TYHTMNSVPC TSTSTVPLAH  
101 HHHHHHHHQA LEPGLDLHI SPSLALMAG AGAGAAGG GGAHDPGCG  
151 GPGGGGGGPG GGGPGGGGGG GPGGGGGGPG GGLGGSANP HPHMHLGHL  
201 SHPAAANAN MPGLPHPL VAAAHNGAA AAAAASAGQ VAAASAAAV  
251 VQAAGLASIC DSDTRPRELE AFAERKORR IKLGVTQADV GSAALAKLP  
301 GVGSLSQSTI CRFSLTSLH NNNIALKPL QAWLEDEGA QRRKMKPEL  
351 ENGGEKKRRR TSIAPERRS LEAYFAVOPR PSSEKIAAIA EKLDLKNV  
401 RWFCNOROK QRRMKFSATY

!!AA\_SEQUENCE 1.0  
P1:G02321 - thyroid transcription factor 1 - human  
C:Species: Homo sapiens (man)  
C>Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 24-Sep-1999  
C:Accession: G02321; G02039  
R:Hamdan, H.  
submitted to the EMBL Data Library, December 1995  
A:Reference number: H01047  
A:Accession: G02321  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-401 <HAM>  
A:Cross-references: EMBL:U43203; NID:91199864; PIDN:AAA89066.1; PID:91199865  
R:Hamdan, H.  
submitted to the EMBL Data Library, August 1995  
A:Reference number: G02039  
A:Accession: G02039  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-71 <HA2>  
A:Cross-references: EMBL:U33627; NID:91113816; PID:91000129  
C:Genetics: TTF-1  
A:Gene: TTF-1  
A:Introns: 26/2  
C:Superfamily: unassigned homeobox proteins; homeobox homology  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:192-248/Domain: homeobox homology <HOX>  
G02321 Length: 401 February 11, 2000 15:51 Type: P Check: 1141 ..



```

1 MWSGGSGKAR GWEAAGAGRS SPEGRSRRRI MSMSPKHTTP FSVSDLSLPL
51 EESYKRVGME GGLGAPLAA YROGAAPPT AAMQOHAVGH HCAVTAIYIM
101 TAAGVPLSH SAVGYCNGN LGMSLELPY QDTRNSASG PGWYCANPDP
151 RFPALSRFG PASGMNSGM GGLSGLDVS KMAPLDSAP RRRRYLFSQ
201 AQYELERRF KOQKYLSPF REHLASMIH LPTQYKIWFQ NHRYYKMRQA
251 KDAQAQOOLQ QDSGGGGGGG GTGCPQOQQA QOOSPRYAV PVLKDGKPC
301 QAGAPAPGAA SLGGHAQQA QHOQAQAAA AAATSVSGSG AGLGARHGQ
351 PGSAGSPDL AHNAAPAL OGQVSSLSHL NSSGSDYGTM SCSTLLYGR
401 W

11AA_SEQUENCE 1.0
P1:S78452 - POU domain protein rdc-1 - human
C:Species: Homo sapiens (man)
C>Date: 19-Mar-1998 #sequence_revision 19-Mar-1998 #text_change 24-Sep-1999
C:Accession: S78452; S26063; I38151
R:Alt, F.W.
Submitted to the EMBL Data Library, February 1992
A:Reference number: S78452
A:Accession: S78452
A:Molecule type: mRNA
A:Residues: 1-331 <ALT>
A:Cross-references: EMBL:X64624; NID:935914; PIDN:CAA5907.1; PID:935915
R:Collum, R.G.; Fisher, P.E.; Datta, M.; Hells, S.; Thiele, C.; Huebner, K.;
Croce, C.M.; Israel, M.A.; Thell, T.; Moroy, T.; Depinho, R.; Alt, F.W.
Nucleic Acids Res. 20, 4919-4925, 1992
A:Title: A novel POU homeodomain gene specifically expressed in cells of the
developing mammalian nervous system.
A:Reference number: I38151; MUID:93027214
A:Accession: S26063
A:Molecule type: mRNA
A:Residues: 174-177, 'S', '179-233', 'K', '235-327' <COL>
A:Cross-references: EMBL:X64624
A:Experimental source: Placenta
C:Superfamily: unassigned homeobox proteins; homeobox homology; POU domain
homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
P1:181-250/Domain: POU domain homology <POU>
F:268-324/Domain: homeobox homology <Hox>
S78452 Length: 331 February 11, 2000 15:51 Type: P Check: 3745 ..

1 MNSVPCHTST VPLAHNNHH HHKQLEPQD LLDHISPSL ALMAGAGARR
51 GAGGGGAHDA AGGGGPRGG GPGGGGGGG GGGGAGGGG GPGGGLLGA
101 SAHPHPMHS LGHLSPHAA AAMNPSGLP HPGIVAAAHA HAAAAAATAA
151 AAGVAAASA AAVVAGAGLA STCSDSTDP ELEAFSGFK RIKIGYVQA
201 DVGSALANIK IPGVSLQS TICRESLTL SHNMIALKP IIOANLEAE
251 GPSEKMPRE LFNGEKKRK RTSIAPEKR SLEAYFAVOP RPSEKIAAI
301 AEKLDLRNV VRWFQNRQ KQRKFSAT Y

11AA_SEQUENCE 1.0
P1:S7329 - probable phosphoesterase (EC 3.1.-.-) - Synecocystis sp. (strain
PCC 6803)
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C>Date: 35-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 24-Sep-1999
C:Accession: S76329
R:Ranko, T.; Saito, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.;
Miyajima, N.; Hirose, M.; Sugitara, M.; Sasamoto, S.; Kimura, T.; Hosouchi,
T.; Matsuno, A.; Mureki, A.; Nakazaki, N.; Natsu, K.; Okumura, S.; Shimo, S.;
Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tanaka, S.
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium
Synecocystis sp. PCC6803. II. Sequence determination of the entire genome and
assignment of potential protein-coding regions.
A:Reference number: S74322; MUID:97061201
A:Accession: S76329
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-416 <KAN>
A:Cross-references: EMBL:D64000; GB:AE001339; NID:91001484; PIDN:BA010181.1;
PID:9101832; PID:91001554
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June
1996
C:Comment: This sequence has motifs characteristic of a variety of
phosphoesterases.
C:Superfamily: unassigned probable phosphoesterases; phosphoesterase core
homology
C:Keywords: hydrolase
F:3-95/Domain: phosphoesterase core homology <PEC>
S76329 Length: 416 February 11, 2000 15:51 Type: P Check: 4655 ..

1 MLKYLHLSDI HUGSGSGSHG IMPATGLNTR LEDFQSLRL CIDRAIAEPA
51 DVVLEGGDAP PDATPPYVQ EAFAPERRL ADADIPYVL VGNHDSGSG
101 SGASLCTYR TLAVPGFVG DRLATHLIFT ANGDIQIVTL PWLIRSTLLT
151 RPEEGSLSE AINLLKLKL QPILGEIRS LDPQPTIIL AHLMADRAE
201 GAETLSGK GETIPALIN RPEEDYVALG HVKHQNINP HNPPIVPG
251 SIERNDEGE KEDGYIWLK ISKGVDMOF CPLPARPFT IKVDTENTD
301 PGGELMAIA KPAIDNAVR LYOIRSEOL EQIDNKLOE ALKPSHSYTI
351 RPELVQAR PRAPELGVGS ALDPLVALKT YLENKDKIKE LMPDIMEAE
401 HLNLSSEQW LEAVGD

11AA_SEQUENCE 1.0
P1:F71508 - probable phosphoesterase (EC 3.1.-.-) CT488 - Chlamydia trachomatis
(sero type D, strain UW3/Cx)
C:Species: Chlamydia trachomatis
C>Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 24-Sep-1999
C:Accession: F71508
R:Stephens, R.S.; Kaiman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.;
Mitchell, W.P.; Olinger, L.; Tatusov, R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans:
Chlamydia trachomatis.
A:Reference number: A71570; MUID:99000809
A:Accession: F71508
A:Molecule type: DNA
A:Residues: 1-244 <ARN>
A:Cross-references: GB:AE001322; GB:AE001273; NID:93328916; PIDN:ACG8088.1;
PID:93328924
A:Experimental source: serotype D, strain UW-3/Cx
C:Comment: This sequence has motifs characteristic of a variety of
phosphoesterases.
C:Genetics:
A:Gene: CT488
C:Superfamily: unassigned probable phosphoesterases; phosphoesterase core
homology
C:Keywords: hydrolase
F:2-85/Domain: phosphoesterase core homology <PEC>
F71508 Length: 244 February 11, 2000 15:51 Type: P Check: 6469 ..

1 MRIFALADH LSLGVPKTM EVRGPWGY HQIKEMRD IVSSDVIOL

```

51 PGDISMAMRL EEAQVDFRFL GALPGIKYMI RGNHDYSSA SSAKLANVLP  
101 ETLHLSKGY VLLNAHOAIV GVALMDSOI CLHMETQHDG PQRYLFEODD  
151 KIFLEKYGRL ERALKELPAS VEDVLVMTYH PVSNDGTPG RVSNTLEMDG  
201 RVSRLCFGL HKVPRPEPGF GNIRGIEYTL VAADYVDIP QVVS

!!AA\_SEQUENCE 1.0  
P1:S35782 - serine/threonine protein kinase - bovine herpesvirus 1  
C:Species: bovine herpesvirus 1  
C:Date: 09-Jun-1994 #sequence\_revision 12-May-1995 #text\_change 24-Sep-1999  
C:Accession: S35782  
R:Audonnet, J.  
Submitted to the EMBL Data Library, June 1993  
A:Reference number: S35782  
A:Accession: S35782  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-467 <AUD>  
C:Cross-references: EMBL:Z23068; NID:9312185; PIDN:CAA80602.1; PID:9312186  
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homology  
C:Keywords: ATP  
F:162-439/Domain: protein kinase homology <KIN>

S35782 Length: 467 February 11, 2000 15:51 Type: P Check: 5295 ..

1 MERAERLAR QRAQLWRSR FACQVAERS GSRLGQSVRG AAAAPARCAA  
51 EGSADLYLAV NNEPEVAPP ARGPPDADG IEGGAAGN EGGVAAAGNE  
101 RAAAGDEKE SASGENESE SESESESE SEADGDMD DDAAGRAGV  
151 TREAEAGAR ALNFRITRL TRESERVE ATGPAPQEH VIKIGASAS  
201 TLAEMLIRT LDHANVVKL AVLFHELVC AVLARYEDL HTLWMIDRP  
251 MALPTALQVT RAVLOGLAYL HSHRIARHV KTEVFLNGP GDVCLDDEGA  
301 AHGPTEPRY YGLAGTLETN SPELLARARY DCRTDWSAG VVAYEMLAYP  
351 RALFSPAGP QGEDAEASGP PTILGDRDCA RQLLRVIRL AVAAEPPPS  
401 PTDRITRNFK RHAATGREPH SPYRCLAVLR LPCDADRLLH QMLTFDFRAR  
451 PTAELLEHP VFGAASG

!!AA\_SEQUENCE 1.0  
P1:CG9230 - conserved hypothetical protein MTH973 - Methanobacterium thermoautotrophicum (strain Delta H)  
C:Species: Methanobacterium thermoautotrophicum  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 24-Sep-1999  
C:Accession: CG9230  
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lum, W.; Pochier, B.; Qiu, D.; Spadafora, R.; Viclir, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso, A.; Bush, D.; Sifer, H.; Patwell, D.; Pribhakar, S.; McDougall, S.; Shimer, G.; Goyal, A.; Pietrokowski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
J. Bacteriol. 179, 7135-7155, 1997  
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functional analysis and comparative genomics.  
A:Reference number: A69000; MUID:98037514  
A:Accession: CG9230  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-257 <MTH>  
A:Cross-references: GB:AE000871; GB:AE000666; NID:92622069; PIDN:AAB85469.1; PID:92622072  
A:Experimental source: strain Delta H  
C:Genetics:

A:Gene: MTH973  
A:Start codon: GTG  
C:Superfamily: hypothetical protein M0915

CG9230 Length: 257 February 11, 2000 15:51 Type: P Check: 3724 ..

1 MIMWVGIV CGAIVNLITG FVRDGRVPYK LGFFYDDLE RAENLMSWD  
51 GRAVLDVADM LPEYDLYVEA ASPEAVRDY PELLEGKV VMSVCLMD  
101 PELREMLVEL ASLNDATIHV PGSAIVGLDG LKASMGNI SVYLITRRPP  
151 RSLGISMDER KYLYRGASE AVKPEPLIN VAAALSLACD RDYDVEITAD  
201 PAVDRNVHEV TVRGDFGEFK TITENVCSV NPKTSVMAAY SAIKLKLSLS  
251 ENIHIGT

!!AA\_SEQUENCE 1.0  
P1:A40369 - trans-acting transcription regulator rbcR - Chromatium vinosum  
C:Species: Chromatium vinosum  
C:Date: 21-Feb-1992 #sequence\_revision 21-Feb-1992 #text\_change 24-Sep-1999  
C:Accession: A40369  
R:Viñale, A.M.; Kobayashi, H.; Akazawa, T.; Henikoff, S.  
J. Bacteriol. 173, 5224-5229, 1991  
A:Title: rbcR (sic, rbcR), a gene coding for a member of the LysR family of transcriptional regulators, is located upstream of the expressed set of ribulose 1,5-bisphosphate carboxylase/oxygenase genes in the photosynthetic bacterium Chromatium vinosum  
A:Reference number: A40369; MUID:91317745  
A:Accession: A40369  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-302 <VIA>  
A:Cross-references: GB:M64032; NID:9145018; PIDN:AAA23327.1; PID:9145020  
C:Superfamily: transcription activator LysR-type  
C:Keywords: DNA binding; transcription regulation

A40369 Length: 302 February 11, 2000 15:51 Type: P Check: 4295 ..

1 MHVSLQRLV FEAVARNISY TRAEELHLS QPAVSMQVRQ LEDELGLSIF  
51 ERLGRQVILT EAGREVFHYS RAIGOSLREM EEVEESLKGV SROSLRIAVA  
101 SVNYFAERL MAIFQSHSG IGLRLDTNR ESLVQMDSN SYDVLVAGVP  
151 PRNVEVEA FMDNPLVIA PPDHPLAGER AISLARLAEE TFMREESGSG  
201 TRQANERFES ERGQTRHGM QMTRNEAVKO AVRSGLGLSV VSLHTELEL  
251 ETRRLVITD EGFPPDRQWY LYRRGKRSL PAAGAFREYV LSRANMHCR  
301 LG

!!AA\_SEQUENCE 1.0  
P1:A33960 - cytochrome caa3 oxidase ctaA - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 23-Mar-1990 #sequence\_revision 18-Sep-1992 #text\_change 24-Sep-1999  
C:Accession: A33960; B66069  
R:Muehler, J.P.; Taber, H.W.  
J. Bacteriol. 171, 4967-4978, 1989  
A:Title: Isolation and sequence of ctaA, a gene required for cytochrome aa-3 biosynthesis and sporulation in Bacillus subtilis.  
A:Reference number: A33960; MUID:89359135  
A:Accession: A33960  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-506 <MUE>  
A:Cross-references: GB:M23915; NID:9142716; PIDN:AAA50254.1; PID:9142717  
A:Note: The authors translated the codon ACT for residue 236 as TTP  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessières, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;

Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conerton, I.F.; Cummings, N.J.; Danile, R.A.; Denizot, F.; Devine, K.M.; Duesterhoft, A.; Ehrlich, S.D.; Emmerston, P.T.; Entlan, K.D.; Erington, J.; Fabre, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A:Accession: 139819  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-18 'A', 20-101 'S', 103-170 'A', 172-188 'A', 190 'F', 192-205 'V', 207-214 'V', 216-252 'G', 254-274 'A', 276-279 'T', 281-288 'T', 290-297 'C' (CUI2)  
 A:Cross-references: GB:M94110; NID:9142782; PIDN:AAA2362.1; PID:9142784  
 A:Experimental source: strain OF4  
 C:Genetics:  
 A:Gene: ctaA  
 A:Start codon: GTG  
 C:Superfamily: cytochrome caa3 oxidase ctaA  
 A:Accession: A45335  
 Length: 297 February 11, 2000 15:51 Type: P Check: 368 ..  
 1 MRRUKIYV ITSIGVILV LQALVTKTG SGEBCATWP LCFGEVTPN  
 51 PAETIIEYS HRIYGLAGA MIIILAWM KQKMBEAK ALSPAVILI  
 101 IFQGLGAGA AVFGSKAIL ALHFGISAMS LAAYVILITL AFDGREHTM  
 151 APRVSKFEY YVEFVITCY GVYSGAYVK HSEATLACDG LPLNGQILFP  
 201 GLYGGGANY FHRVAGTILL LFLILIMTW LSRVRRVVL TWAVLSFLL  
 251 VVAQFISGIS IVFQNALSV GIHILITISK LFSALSYAM IITRPSH  
 11AA SEQUENCE 1.0  
 P1:G02514 - P2 purinoceptor - human  
 C:Species: Homo sapiens (man)  
 C>Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 24-Sep-1999  
 C:Accession: G02514  
 R:Hammett, F.; Southey, M.C.; Somers, G.R.; Hutchins, A.M.; Venter, D.J., submitted to the EMBL Data Library, March 1996  
 A:Reference number: H01373  
 A:Accession: G02514  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-326 'C' (CUI2)  
 A:Cross-references: EMBL:U52464; NID:91407632; PIDN:AB03572.1; PID:91407633  
 C:Genetics:  
 A:Gene: P2Y6  
 C:Superfamily: ATP receptor P2u  
 G02514 Length: 326 February 11, 2000 15:51 Type: P Check: 2327 ..  
 1 MDNGTGALG LPPTCYRE NFKQILIPV YSAVLACLP LNICVITQIC  
 51 TSRRLTRTA VYTNLALAD LLYAGSLPIL IYNAQGDHW PEGDFACRLV  
 101 RFLVYANLNG SILFLTCSF QRYLGICHP LPPHRRGGR AAWLCVAVW  
 151 LAVTQCLPT AIFATGIGR NRTVCYDLS PALATHVMPY GMAITVIGFL  
 201 LPFAALLACY CILACRLCQ DGPAPPAOE RRGKAARNAV VAAAFATSF  
 251 LPFHITKAY LAVRSTPGV CTVLEAFAPA YKTRPFASA NSVLDEILEY  
 301 FTQKFRFRP HELLQKTFK WROGR  
 11AA SEQUENCE 1.0  
 P1:JC4800 - P2Y6 receptor - human  
 C:Species: Homo sapiens (man)  
 C>Date: 15-Oct-1995 #sequence\_revision 16-Aug-1996 #text\_change 04-Sep-1998  
 C:Accession: JC4800  
 R:Comun, A.; Parmentier, M.; Boeynaems, J.M., Biochem. Biophys. Res Commun. 222, 303-308, 1996  
 A:Title: Cloning, functional expression and tissue distribution of the human P2Y6 receptor.  
 A:Reference number: JC4800; MUID:96222498  
 A:Molecule type: cDNA  
 A:Residues: 1-328 'C' (CUI2)  
 A:Note: sequence extracted from NCBI backbone (NCBI:121483); this sequence is the sequence shown in Fig. 2  
 A33960 Length: 306 February 11, 2000 15:51 Type: P Check: 1522 ..  
 1 MNKALKALGV LTFVMIYVL IGGALVTKTG SGOGCGRMP LCHGRFFPEL  
 51 NPASIEESH RFASGISIIL VLSLAFWSMR KIPPIRETT FLAISIIFL  
 101 FLQALGALA VFGSNALIM ALHFGISLIS FASVILITIL IFEADKSVRT  
 151 LKPLQIGCK MGFHIGIIL YSYIVYTGA YVRHTESSLA CPNVPLCSPL  
 201 NNGLPQFHE WQMGHRAAA LLEFWIIVA AVHATSYKD OKQIFMGWIS  
 251 CLIFITLQAL SGIMIVSEL ALGFALHSF FIACTFGVLC YFLLIARRR  
 301 YESROS  
 11AA SEQUENCE 1.0  
 P1:A45333 - cytochrome-c oxidase assembly factor ctaA - Bacillus firmus  
 C:Species: Bacillus firmus  
 C>Date: 30-Apr-1993 #sequence\_revision 18-Nov-1994 #text\_change 24-Sep-1999  
 C:Accession: A45333; I39819  
 R:Quirk, P.G.; Hicks, D.B.; Krulwich, T.A., J. Biol. Chem. 268, 678-685, 1993  
 A:Title: Cloning of the cta operon from alkaliphilic Bacillus firmus OF4 and characterization of the pH-regulated cytochrome caa3 oxidase it encodes.  
 A:Reference number: A45335; MUID:93107080  
 A:Accession: A45333  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 1-297 'C' (CUI2)  
 A:Note: sequence extracted from NCBI backbone (NCBI:121483); this sequence is the sequence shown in Fig. 2

A:Cross-references: EMBL:X57058  
 A:Experimental source: Placenta  
 C:Superfamily: ATP receptor P2u  
 C:Keywords: glycoprotein; placenta; receptor; transmembrane protein  
 F:26-52/Domain: transmembrane #status predicted <TM1>  
 F:63-86/Domain: transmembrane #status predicted <TM2>  
 F:104-122/Domain: transmembrane #status predicted <TM3>  
 F:143-167/Domain: transmembrane #status predicted <TM4>  
 F:193-216/Domain: transmembrane #status predicted <TM5>  
 F:241-264/Domain: transmembrane #status predicted <TM6>  
 F:283-305/Domain: transmembrane #status predicted <TM7>  
 F:5.173/Binding site: carbohydrate (Asn) (covalent) #status predicted

Jc4800 Length: 328 February 11, 2000 15:51 Type: P Check: 7545 ..

1 MEMONGGQA LGPPTCY RKNRQLLP PYSAVLAAG LPLNCVITO  
 51 ICTRRALTR TAVYTLNLA ADLYACSLP LLTYNAQSD HMPEDFACR  
 101 LVRLFLPANT HGSILFLTCI SFQRYLGICH PLAPWHRKG RRAANLVCA  
 151 VWLAATTOCL PRAIFAATGI QRNRTVCYL SPALATHYM PYGMALTVIG  
 201 FLPLFAALLA CYCLACRLC RODGPAPVA QERRGRAAM AVVAAPAI  
 251 SLEPHITKT AYLAVRTPG VPCVLEAFA AAYKTRPFA SANSLDPIL  
 301 FYFOKKRR RPHELOKLT AKWROGR

!!A\_SEQUENCE 1.0  
 P1:P64471 - hypothetical protein homolog MJ1375 - Methanococcus jannaschii  
 C:Species: Methanococcus jannaschii  
 C>Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 04-Sep-1998  
 C:Accession: F64471  
 R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.C.; Blake, J.A.; Fitzgerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghegan, N.S.N.; Melman, J.E.; Fuhrman, J.L.; Nguyen, D.; Venter, J.D.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hrust, M.A.  
 Science 273, 1058-1073, 1996  
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.  
 A:Reference number: A64300; MUID:96337999  
 A:Accession: F64471  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-415 <BUL>  
 A:Cross-references: GB:U67577; GB:L77117; NID:g1592013; PID:g1592020; TIGR:MJ1375; PID:g1511391  
 C:Genetics:  
 A:Map position: REV1324199-1322952  
 C:Superfamily: hypothetical protein MJ1375

F64471 Length: 415 February 11, 2000 15:51 Type: P Check: 3696 ..

1 MSLCKDSIYI LMSNLYSKM AYLFYITAF LLGTEAFGI KGLMPIDTL  
 51 TIFSSGIPP ALAKFLAEK EVDINKYPI LYMLILSV GFILPYIKY  
 101 ILGGHYLNP NLTYAVGLC VVASTVAFS RGILOGLKN KYLSLTWIV  
 151 YTKAVILVFI LTLYLIGIFS LLSISLAVLV GGIFGYLLIY KALKKGFEPK  
 201 KLIDIKWTK NIFSNENLDI LRSPIALT SSSYRLEGGI DNIVMSING  
 251 GFMGSIYGS SXISRGIMF ASAVISILP RISKTDLSI LKEGIQNTI  
 301 FSSIFVIGCL FPEPIPLAF EKTANPEIL CLRIALSSI FMSYTLISS

351 ALDGLGYAKI SFYILFGLV LNTILNLYV NAYGVGSL ATLITISIV  
 401 LIVERFALRI KKHII

!!A\_SEQUENCE 1.0  
 P1:A47201 - bombesinlike peptide - African clawed frog  
 C:Species: Xenopus laevis (African clawed frog)  
 C>Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 04-Sep-1998  
 C:Accession: A47201  
 R:Wachselberger, C.; Krell, G.; Richter, K.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 9819-9822, 1992  
 A:Title: Isolation and sequence of a cDNA encoding the precursor of a bombesinlike peptide from brain and early embryos of Xenopus laevis.  
 A:Reference number: A47201; MUID:93028554  
 A:Accession: A47201  
 A>Status: preliminary  
 A:Molecule type: nucleic acid  
 A:Residues: 1-120 <WEC>  
 A>Note: sequence inconsistent with the nucleotide translation  
 A>Note: sequence extracted from NCBI backbone (NCBI:115857, NCBI:115858)  
 C:Superfamily: ranatensis

A47201 Length: 120 February 11, 2000 15:51 Type: P Check: 1219 ..

1 MSAPVLRML PLRFLTLL LSPFLYFCM ESEDAKRIE KIRGNOWAI  
 51 GFMGKKSIO DTYNPEODM DSEDFRPRI EMIAGTFROE PIRALSPKQ  
 101 DEIOMMLKRI MDDYIKITQK

!!A\_SEQUENCE 1.0  
 P1:E65139 - hypothetical protein b3434 - Escherichia coli (strain K-12)  
 C:Species: Escherichia coli  
 C>Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 24-Sep-1999  
 C:Accession: E65139  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:97426617  
 A:Accession: E65139  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-197 <BLAT>  
 A:Cross-references: GB:AE000420; GB:U00096; NID:g1789840; PIDN:AAC76459.1; PID:g1789842; UMG:p.b3434  
 A:Experimental source: strain K-12, substrain MG1655  
 C:Superfamily: conserved hypothetical protein MJ1677

E65139 Length: 197 February 11, 2000 15:51 Type: P Check: 1678 ..

1 MNEIISAAYL LILMDPLGN LPIFMSVLKH TPKRRRAIM VRELLIALV  
 51 MLYFLRGER ILAFLSLRAE TVISGCIIL FLIAKMFIP SASNSSGLP  
 101 AGEPEPIVPL AIPYVAGPTI LATLMLSHQ YQNGHGLVI ALLAMGTF  
 151 VILLOSLFL RLKGEKVNA LERLMGLIV MAAQMFIDG IRMMWKG

!!A\_SEQUENCE 1.0  
 P1:S48737 - kynurenine aminotransferase - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 07-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 26-Aug-1999  
 C:Accession: S48737  
 R:Mosca, M.; Cozzi, L.; Breton, J.; Speciale, C.; Okuno, E.; Schwarcz, R.; Benatti, L.  
 FEBS Lett. 353, 21-24, 1994  
 A:Title: Molecular cloning of rat kynurenine aminotransferase: identity with glutamine transaminase K  
 A:Reference number: S48737; MUID:95010756  
 A:Accession: S48737  
 A>Status: preliminary

A:Molecule type: mRNA  
 A:Residues: 1-423 <MOS>  
 A:Cross-references: GB:S74029; NID:9688257; PID:9688258  
 C:Superfamily: aspartate transaminase

S48737 Length: 423 February 11, 2000 15:51 Type: P Check: 6122

1 MTRKLAQRRL DGDONLWVE FGKLTKEYDV VNLGGFDPF SPPDFATQAF  
 51 QQAATNGFNL NQYTRAFGP PLTVNLASF GKLLGQEMP LTNLVTVGA  
 101 YGALFTAFQA LVDEGEVIL MEPAFDCYEP MTMAGGCPV FVTLKPSAP  
 151 KGRKGAANDM QDPAELASK FTPTKVLV NTPNNPLGKV FSRMELEVA  
 201 NLCOQHDVVC ISDEVYQWLV YDGHQVISA SLPGMMDRL TIGSAGKSF  
 251 ATGKRVGWM GPDNIMKHLR TVHONSIFHC PTOQAANAQ CFERQOQHF  
 301 QPSSTFLQLP QAMELNDRHM IRSLSQVGLK LMSQGSYPL IADISDFSK  
 351 MPDLPGADE PYDRRAKMA IKNGLVGIP VSTFFSRPHO KDFHVIKFC  
 401 FVKDKATLQA MDERLRKKE LQP

!IAA\_SEQUENCE 1.0  
 P:J02268 - O-methyltransferase (EC 2.1.1.-) - maize  
 C:Species: Zea mays (maize)  
 C>Date: 28-Aug-1985 #sequence\_revision 07-Oct-1994 #text\_change 29-Sep-1999  
 C:Accession: J02268  
 R:Held, B.M.; Wang, H.; John, I.; Wurtele, E.S.; Colbert, J.T.  
 Plant Physiol. 102, 1001-1008, 1993  
 A:Title: An mRNA putatively coding for an O-methyltransferase accumulates preferentially in maize roots and is located predominantly in the region of the endodermis.  
 A:Reference number: J02268; MUID:94105316  
 A:Accession: J02268  
 A:Molecule type: mRNA  
 A:Residues: 1-364 <HEL>  
 A:Cross-references: GB:I14063; NID:9404069; PID:MAA18532.1; PID:9404070  
 A:Experimental source: root, cv. NKH31  
 C:Superfamily: O-methyltransferase  
 C:Keywords: methyltransferase

J02268 Length: 364 February 11, 2000 15:51 Type: P Check: 6645

1 MELSPNNSTD OSLDAQLEL WTTTFPFMS MALSAIHLR IADAIHLHG  
 51 AASLSQILSK VHLHPSRVSS LRRLMRYLFT TNVFGTOLG GGSDDSEFY  
 101 YLTPVSRLLT IGSOSQOLAQ TPIANVLDP TIVSPFSLG AMFQHLPPD  
 151 CLFKHTHGRG IWELEKDDAT FDALVNDGLA SDSQILYDVA IKOSAEVFG  
 201 ISSLVGVGGG IGAAQAALSK APFHVCSVL DLAHVAKAP THTDVQFLG  
 251 DMFESIPPAD AVLLKSVLHD WDHDQVKIL KNCKRAIPR EAGGKVIIN  
 301 MVVGAQPSDM KKEKQALFD VYIMFINGME RDEDEMSKIF SEAGISDYRI  
 351 IPIVGVRSII EYYP

!IAA\_SEQUENCE 1.0  
 P:G70165 - aldose reductase homolog - Lyme disease spirochete  
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
 C>Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 26-Aug-1999  
 C:Accession: G70165  
 R:Fraser, C.M.; Castjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.; Roddson, R.; Hickey, E.K.; Gwin, M.; Dougherty, B.; Tomb, J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman, J.; Utterback, L.; Matthews, L.; McDonald, L.; Artiach, P.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.;

Robert, K.; Hatch, B.  
 Nature 390, 580-586, 1997  
 A:Authors: Smith, H.O.; Venter, J.C.  
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.  
 A:Reference number: A70100; MUID:98065943  
 A:Accession: G70165  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-315 <HEL>  
 A:Cross-references: GB:AE001154; GB:AE000783; NID:92688431; PID:92688432;  
 TIGR:B80528  
 C:Superfamily: conserved hypothetical protein YP1088w

G70165 Length: 315 February 11, 2000 15:51 Type: P Check: 3950

1 MNMLDKINT YSKLLIGSQO FGGGYKQVE KETAKKILK AYDGIKINID  
 51 TARAYNGIS EKIIGELIER DPTIRENLI ASKCYMEIS EYENFNEISL  
 101 KNLKTDYIDI YYIHWKADF DLRIPIASFE EMKVKGRIRY VGSNFEISH  
 151 MESIKRVCKI DVNOIGYNPL FRNKEKDVIP YCEDNNIATI SYSTIAOGL  
 201 SKANIKDKNK FNDIREKLI LFKKEIWPYT LKTINKLEI AKINMLTLE  
 251 LYSWLKRTK LSGPIVFSK ENYVESNVS FKAELNDKY EITISILDNF  
 301 NHQTNFPL FNKKI

!IAA\_SEQUENCE 1.0  
 P:G69076 - molybdopter in biosynthesis protein Moeb homolog - Methanobacterium thermoautotrophicum (strain Delta H)  
 C:Species: Methanobacterium thermoautotrophicum  
 C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Sep-1999  
 C:Accession: G69076  
 R:Smith, D.R.; Doucette-Stamm, L.A.; Delonghery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Basillzaden, K.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lum, M.; Pochier, B.; Qiu, D.; Spedding, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibso, J.; Jiwan, N.; Caruso, A.; Bush, D.; Sater, H.; Patwell, D.; Prabhakar, S.; McDougall, S.; Shiner, G.; Goyal, A.; Pietrovski, S.; Church, G.M.; Daniels, C.U.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
 J. Bacteriol. 179, 7135-7155, 1997  
 A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functional analysis and comparative genomics.  
 A:Reference number: A69000; MUID:98037514  
 A:Accession: G69076  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-251 <MTH>  
 A:Cross-references: GB:AE000917; GB:AE000666; NID:92622689; PID:92622694  
 A:Experimental source: strain Delta H  
 C:Genetics:  
 A:Gene: MTH1571  
 C:Superfamily: molybdopter in biosynthesis protein moeb

G69076 Length: 251 February 11, 2000 15:51 Type: P Check: 4836

1 MPERYEMAY WEMVSROMGL LSRADQLKX DSTVSVICG GIGGAEMVL  
 51 ARMVGSLRI IDSDVFDVSN INQMLSSFR DLRIKVDVA AERIRIVNPE  
 101 SRKVYHEIF DEENAEIIP GSDAVVDALD NITSKVIAS RCHSEGIPFI  
 151 HGAHSGMGQ VTFVREGSPS YEELFRLPSV GELTADYKA KLRDLSKTP  
 201 PVIGPVANIT GGLQAAEVLK LITGRGDVIA APRMLKFDLF OGEPFKIVEL  
 251 S

!IAA\_SEQUENCE 1.0  
 P:J56490 - hypothetical transcription regulator pepa-gntv intergenic region -

Escherichia coli  
 M:Alternate names: hypothetical protein f332  
 C:Species: Escherichia coli  
 C:Date: 28-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 29-Sep-1999  
 C:Accession: S56490; G65239  
 R:Burkhardt, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.  
 Nucleic Acids Res. 23, 2105-2119, 1995  
 A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region  
 from 92.8 through 100 minutes.  
 A:Reference number: S56314; MUID:95334362  
 A:Accession: S56490  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-332 <BUR>  
 A:Cross-references: EMBL:U14003; NID:91263172; PID:AAA97161.1; PID:g537106  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;  
 Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,  
 J.; Davis, N.M.; Kapatnick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:97426617  
 A:Accession: G65239  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-332 <BLAT>  
 A:Cross-references: GB:AE000497; GB:000096; NID:91790711; PID:AACT7221.1;  
 PID:g190715; UMEP:D4264  
 A:Experimental source: strain K-12, substrain MG1655  
 C:Genetics:  
 A:Gene: yjg5  
 C:Superfamily: gntR protein  
 S56490 Length: 332 February 11, 2000 15:51 Type: P Check: 2843 ..

1 MRNRHLEOD IATLACVTM TVSRYSRSPK KVARETERI AKIMEINYI  
 51 PNRAFGMLN AOSYTLGILI PSFONQLFAD ILAGIESVTS ENHYQQLIN  
 101 YNDRDSEEE SVINILSYNI DGIILSEKYH TIRTVKRLS AIIPEVLEMD  
 151 VQGERLMEV GEDNRQAFD MCVTLERKVR RHKLIVGSK DDTREQRKQ  
 201 GYCDAMMLN LSPILRNPRP ISSIHLGMQL MRDALSNDP LDCVCTND  
 251 IAWGALLCR ERNLAVPEI SIAGHGLEI GRQMPSLAS VTPREDFGR  
 301 MAQMLLSKI KNNHNHNTV DLGQIYHGN TL

11AA:SEQUENCE 1.0  
 P1:H69823 - conserved hypothetical protein yHt - Bacillus subtilis  
 C:Species: Bacillus subtilis  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 29-Sep-1999  
 C:Accession: H69823  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo,  
 V.; Bertero, M.G.; Bessières, P.; Bolotin, A.; Borchert, S.; Boriss, R.;  
 Boursier, L.; Brans, A.; Braun, M.; Brignelli, S.C.; Bron, S.; Broutillet, S.;  
 Buschli, C.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani,  
 J.J.; Conerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denicourt, F.; Devine, K.M.;  
 Dueselherdt, A.; Ehrlich, S.D.; Emerson, P.T.; Ertlan, K.D.; Errington, J.;  
 Feibert, C.; Ferrati, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi,  
 A.; Gallier, N.; Gilm, S.Y.; Glasner, J.; Goffeau, A.; Goldschmidt, E.J.; Grandt,  
 G.; Guillep, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.;  
 Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.;  
 Joris, B.; Karamata, D.; Kasahara, Y.; Kjaer-Blaehard, M.; Klein, C.;  
 Kobayashi, Y.; Koestler, S.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.;  
 Lapilus, A.; Lardinois, S.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda,  
 S.; Mausel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.;  
 Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega,

B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,  
 A.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds,  
 S.; Rieger, M.; Rivolta, C.; Kocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato,  
 T.; Scanlon, E.  
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska,  
 A.; Serot, S.J.; Serrot, P.; Shin, B.S.; Solido, B.; Sorokin, A.; Tacconi, E.;  
 Takagi, I.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Takakoshi, A.; Tanaka,  
 T.; Terpe, P.; Tothoni, A.; Tosato, V.; Uchiyama, S.; Vandembol, M.;  
 Vannier, F.; Vassaret, A.; Vairi, A.; Wambutt, R.; Wedler, E.; Wedler, H.;  
 Weitzenger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto,  
 K.; Yata, K.; Yoshida, K.  
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus  
 subtilis.  
 A:Reference number: A69580; MUID:98044033  
 A:Accession: H69823  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-302 <KUN>  
 A:Cross-references: GB:Z99108; GB:AL009126; NID:92633055; PID:CA12749.1;  
 PID:el182910; PID:g2633244  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: yHt  
 C:Superfamily: conserved hypothetical protein H10176  
 H69823 Length: 302 February 11, 2000 15:51 Type: P Check: 1929 ..

1 MNQKRGLEI LNKQDQGW LSVLKTALK ASKPYIDDM SHQIYVNE  
 51 SVLNMYKX GDRVFDLQ SEASSVPEY GELDILEDN HMLINKPAG  
 101 IATHNEDGO TGLANLAI HQINGETCK VRHVRLDQD TSGAIVFAH  
 151 RLAAHLDQO LEKTLKRYT TAIAGKLRT KKGITNPIC RDRSHPTRR  
 201 VSPGGQIVT HFVYASNAK ERLSVELEL ETGRTHQIRV HLASLGHPT  
 251 GDSLYGGSK LNRQALHAN KVQAVHPITD ELIYVAFPP ADMKNLCRTY  
 301 FS

11AA:SEQUENCE 1.0  
 P1:A64223 - probable pseudouridine synthase - Mycoplasma genitalium (SGC3)  
 C:Species: Mycoplasma genitalium  
 C:Date: 10-Nov-1995 #sequence\_revision 10-Nov-1995 #text\_change 26-Aug-1999  
 C:Accession: A64223; T09718  
 R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.;  
 Fleischmann, R.D.; Bult, C.J.; Kerlavage, A.R.; Sutton, G.; Kelley, J.M.;  
 Fritchman, J.L.; Weidman, J.F.; Small, K.V.; Sandusky, M.; Fuhmann, J.;  
 Nguyen, D.; Utlarback, T.R.; Sander, D.M.; Phillips, C.A.; Merrick, J.M.; Tomb,  
 J.F.; Dougherty, B.A.; Bott, K.F.; Hu, P.C.; Lucier, T.S.; Peterson, S.N.;  
 Smith, H.O.; Hutchison III, C.A.; Venter, J.C.  
 Science 270, 397-403, 1995  
 A:Title: The minimal gene complement of Mycoplasma genitalium.  
 A:Reference number: A64200; MUID:96026346  
 A:Accession: A64223  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1308 <TIGR>  
 A:Cross-references: GB:U039698; GB:U03967; NID:91045891; PID:91045895; TIGR:MG209  
 A:Experimental source: strain G-37  
 R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.;  
 Fleischmann, R.D.; Bult, C.J.; Kerlavage, A.R.; Sutton, G.G.; Kelley, J.M.;  
 Fritchman, J.L.; Weidman, J.F.; Small, K.V.; Sandusky, M.; Fuhmann, J.L.;  
 Nguyen, D.T.; Utlarback, T.; Sander, D.M.; Phillips, C.A.; Merrick, J.M.; Tomb,  
 J.; Dougherty, B.A.; Bott, K.F.; Hu, P.C.; Lucier, T.S.; Peterson, S.N.; Smith,  
 H.O.; Venter, J.C.  
 submitted to the EMBL Data Library, October 1998  
 A:Reference number: 216818  
 A:Accession: T09718  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA

A:Residues: 1-308 <FRA>  
 A:Cross-references: EMBL:U39700; NID:93844800; PID:93844807  
 A:Experimental source: Isolate G37  
 C:Genetics: M6209  
 A:Gene: M6209  
 A:Genetic code: SGC3  
 C:Superfamily: conserved hypothetical protein HI0176  
 A64223 Length: 308 February 11, 2000 15:51 Type: P Check: 590 ..

1 MROCFVYVTT KRDSLSL LNSRYKVK LNMQIKVN EKLTFKNSLI  
 51 VARDVYKVE IHDETSDFI TSEVPYKAL EYLFEDKRLM VINKPSGLLT  
 101 HPTTFNEKAS LLAACIFHNN KNPIYLVHRL DDTSGALIV CKNOASLNL  
 151 QNOLNRTLK RYVALVHFP FNALTSINA PLARVNNKV MEKIAQTAKA  
 201 KOATKFKVI NONKKAALS LELLTGRTQ INVHLKTIQH PYINDPLTGI  
 251 KSEKDSYGO FLHNRICFI HPTLNKPMDF HAPLEPKFST KLSLSLSLT  
 301 DPLHLEFK

11AA\_SEQUENCE 1.0  
 PI:573869 - hypothetical protein ycec - Mycoplasma pneumoniae (ATCC 29342)  
 N:Alternate names: hypothetical protein HI0\_crf309  
 C:Species: Mycoplasma pneumoniae  
 A:Date: 27-Feb-1997 #sequence\_revision 25-Apr-1997 #text\_change 26-Aug-1999  
 C:Date: 27-Feb-1997 #sequence\_revision 25-Apr-1997 #text\_change 26-Aug-1999  
 A:Accession: S73869 Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Hermann, R.  
 R:Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Hermann, R.  
 Nucleic Acids Res. 24, 4420-4449, 1996  
 A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae.  
 A:Reference number: S73327; MUID:9710585  
 A:Accession: S73869  
 A:Status: Preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-309 <HIM>  
 A:Cross-references: EMBL:AE000053; GB:U00089; NID:q1674236; PID:q1674240  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996  
 C:Genetics:  
 A:Gene: ycec  
 A:Genetic code: SGC3  
 C:Superfamily: conserved hypothetical protein HI0176  
 S73869 Length: 309 February 11, 2000 15:51 Type: P Check: 6533 ..

1 MEQTFEVTTA QRLDTFLATL LNSRYKVK LIVDGLSVN GKRTKNGWL  
 51 VQPEDVAVHN WSELEFEKVP VEVOYPDPL DILEDEQIM VVKNPGLIS  
 101 HPTSFNESES LLGAALFHCN HOPVFLVHRL DDTSGVIML AKNOSLILH  
 151 QKOLQORVAK RYVALVHFP LDSLSTISA PLERYGNKV MMRVGNSSNK  
 201 AKNAFTKTV LNQNKAALI KCBLLGRTH QIRVHLQRTK HPIYNDPLVG  
 251 LKSDQATEYG QYLAQOISF IHTLNKEMK FEAQDKRFS DKLDNINLKI  
 301 ANSLYALFQ

11AA\_SEQUENCE 1.0  
 PI:B64026 - hypothetical protein HI1338 - Haemophilus influenzae (strain Rd KW20)  
 C:Species: Haemophilus influenzae  
 C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 29-Sep-1999  
 C:Accession: B64026  
 R:Feilschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirnass, E.F.; Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.; Dougherty, B.A.; Merrick, J.M.;

Mckenney, K.; Sutton, G.; FitzHugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton, M.D.; Uterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fritchman, J.L.; Geoghegan, N.S.M.  
 Science 269, 496-512, 1995  
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.  
 A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
 A:Reference number: A64000; MUID:95350630  
 A:Accession: B64026  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-164 <TIGR>  
 A:Cross-references: GB:U32813; GB:I42023; NID:91574796; PIDN:AAC22985.1; PID:91574807; TIGR:HI1338  
 C:Superfamily: hypothetical protein HI1338  
 B64026 Length: 164 February 11, 2000 15:51 Type: P Check: 2974 ..

1 MNIFIMRGE AEVANSRDKA RHLTVGSKQ AFLQGQWIKQ HSTLVINSL  
 51 DRIVSPYR AOETFOVNO APDELENNF EIWEGITPYG HAHVIDYLE  
 101 VLKDEGVKSV LVSHLPLVG EIVALEYGRK NPISFPATI AOLMDGNKS  
 151 EILMHQASPV IYLK

11AA\_SEQUENCE 1.0  
 PI:B65007 - hypothetical protein b2340 - Escherichia coli (strain K-12)  
 C:Species: Escherichia coli  
 C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 29-Sep-1999  
 A:Accession: B65007  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:97426617  
 A:Accession: B65007  
 A:Status: Preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-161 <BLAT>  
 A:Cross-references: GB:AE000322; GB:U00096; NID:91788672; PIDN:AAC75400.1; PID:91788681; UMG:PD2340  
 A:Experimental source: strain K-12, substrain MG1655  
 C:Superfamily: hypothetical protein HI1338  
 B65007 Length: 161 February 11, 2000 15:51 Type: P Check: 9360 ..

1 MGVFIMRHD AALDAASDV RLITNGCDE SRLANWKG OKVEIERLV  
 51 SPFLAEQTL EIVGDCINLP SAEVILPELT PCGDVGLVSA YLQALINQV  
 101 ASYLVISHLP LVGYLVALC PETPPEMTT SALSATLDE SGNTEFNQM  
 151 SPCLMKAKA I

11AA\_SEQUENCE 1.0  
 PI:C64882 - Probable transcription regulator ycjZ - Escherichia coli  
 C:Species: Escherichia coli  
 C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 29-Sep-1999  
 A:Accession: C64882  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:97426617  
 A:Accession: C64882  
 A:Status: Preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA

A:Residues: 1-299 <BLAT>  
 A:Cross-references: GB:AE000231; GB:U00096; NID:g1787388; PIDN:AAC74410.1;  
 PID:g1787589; UMGF:b1328  
 A:Experimental source: strain K-12, substrain MG1655  
 C:Genetics:  
 A:Gene: yjc12  
 C:Superfamily: hypothetical protein b1328  
 C:Keywords: DNA binding; transcription regulation  
 F:20-50/Region: regulatory protein lysR motif

C64882 Length: 299 February 11, 2000 15:51 Type: P Check: 51

1 MKREIADLM AFVVAEERS FTRAAALSM AQSALSQIVR RIERLGLRL  
 51 LRTTRSVAP TEAGEHLLSV LGPMLHDIDS AMASLSDION RPSGTRIRTT  
 101 VEHAKTILL PAMRTFLKSH PEIDIOITLD YGLTDVYSER FDAQVALGGE  
 151 MDKMIATRI GPDIPMAIVG SPDYFSRSRY PTYSQILDH QAINLILPTS  
 201 GTANRWMLIR GGREVRYRME GQLLNTIDL IIDAIDGHG LAYLPDYOE  
 251 RAIREKLIR VLDKFTPDLP GYHLYPHRR HAGSAFSLFT DRLKRYGAV

!!AA\_SEQUENCE 1.0  
 P1:E72058 - Probable phosphatase/kinase - Chlamydia pneumoniae (strain CWL029)  
 C:Species: Chlamydia pneumoniae  
 C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 26-Aug-1999  
 C:Accession: E72058  
 R:Kilman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.;  
 Grimwood, J.; Davis, R.W.; Stephens, R.S.  
 Nature Genet. 21, 385-389, 1999  
 A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
 A:Reference number: A72000; NID:920606  
 A:Accession: E72058  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-202 <RNA>  
 A:Cross-references: GB:AE001645; GB:AE001363; NID:g4376896; PID:g4376907  
 A:Experimental source: strain CWL029  
 C:Genetics:  
 A:Gene: yACE  
 C:Superfamily: conserved hypothetical protein YDR196C  
 E72058 Length: 202 February 11, 2000 15:51 Type: P Check: 4148

1 MRLKVSIT GDLSSGKTEA CQVFOELGAY VVSADDEISHS FLIPHTRIGR  
 51 RVIDLGSIV VVDGAFDAQA IAAKVFYNSV LLOGDEALIH PEVCRIITEQ  
 101 YHOSIDGNY PLFVAEVELL YEIHVAKWPD SVILWMANED IRRERFKKT  
 151 GRSSEDFDR CSRLNVEEK LAQADVVEN NQTKRELKOK IEEFYALKG  
 201 AL

!!AA\_SEQUENCE 1.0  
 P1:B71509 - Probable phosphatase/kinase - Chlamydia trachomatis (serotype D, strain UM3/Cx)  
 C:Species: Chlamydia trachomatis  
 C>Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 26-Aug-1999  
 C:Accession: B71509  
 R:Stephens, R.S.; Kilman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.;  
 Mitchell, W.P.; Olinger, L.; Tatusov, R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.  
 Science 282, 754-759, 1998  
 A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis.  
 A:Reference number: A71570; MUID:99000809  
 A:Accession: B71509  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-202 <RNA>  
 A:Cross-references: GB:AE001322; GB:AE001273; NID:g3328916; PID:g3328928

A:Experimental source: serotype D, strain UM-3/Cx  
 C:Genetics:  
 A:Gene: yACE  
 C:Superfamily: conserved hypothetical protein YDR196C

B71509 Length: 202 February 11, 2000 15:51 Type: P Check: 4232

1 MDLKIYSVT GDSSGKTEA CQVFEDLQAY VISADKVSHS FLVPTSVQ  
 51 RIIDLGPET IIENTLSRKA IAEKVFGRND LLISLEILH PEVCRFVEEK  
 101 YAHVQOEKY PLFIAEPPL YEIOYADWFD QVILISADTG IRRERFLKKT  
 151 GGSIDSEFLR CARFSSLEEK ILRADVIEN NGTKKEFRK VQCFKALKG  
 201 TI

!!AA\_SEQUENCE 1.0  
 P1:D64151 - hypothetical protein H10404 - Haemophilus influenzae (strain Rd KW20)  
 C:Species: Haemophilus influenzae  
 C>Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 29-Sep-1999  
 C:Accession: D64151  
 R:Flieschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.;  
 Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.; Dougherty, B.A.; Merrick, J.M.;  
 McEneaney, K.; Sutton, G.; Fitzhugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;  
 Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton, M.D.; Utterback, T.R.; Hanna, M.C.;  
 Nguyen, D.T.; Saudek, D.M.; Brandon, R.C.; Fine, L.D.; Fitchman, J.L.;  
 Furumasa, J.L.; Geoghegan, N.S.M.  
 Science 269, 496-512, 1995  
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.;  
 Venter, J.C.  
 A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
 A:Reference number: A64000; MUID:95350630  
 A:Accession: D64151  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-430 <TIGR>  
 A:Cross-references: GB:U02273; GB:L42023; NID:g1573363; PIDN:AAC22063.1;  
 PID:g1573375; TIGR:H10404  
 A:Note: best homolog was a hypothetical protein from Bacillus subtilis  
 C:Superfamily: hypothetical protein H10404  
 D64151 Length: 430 February 11, 2000 15:51 Type: P Check: 1410

1 MDLSDIERQ LOKATAQAFI IALSGGLDST VLSLFAKLC QKQPHLPPLS  
 51 IRAIHHLGL SPNADSWAKH CODLQDOFQI PLIERVOYD KTINGIAGAR  
 101 EARYQAIKKY LQTOEMLVTA HHLNDQTEFF FLAKRQSGI KGLGAAQOOS  
 151 VLEGPIILRP LIGFTRTQLE NYAOKKELNW ITDSNEDNR YDRNFIENET  
 201 LPELERNVAH FDLAVQRSNQ HCFEQOQLIN DLSEITTEH COIKNOFKIC  
 251 QFQYSLAKQ TALLRWLAE NQLEMPSKRO LFOILINDYIF AKEANPOFO  
 301 LVNKYIRIYQ DSIXLTKRPS DLTCTKLE QNTLNPDDL GNLTVQENEH  
 351 NLITVWDYS VLEKTNLPI SIRQYSGKV KHYPRPRRD IKKIWOELGV  
 401 PPWERNRPL IFYGNELKSA VGFFRVLKSS

!!AA\_SEQUENCE 1.0  
 P1:S75967 - hypothetical protein - Synechocystis sp. (strain PCC 6803)  
 C:Species: Synechocystis sp.  
 A:Variety: PCC 6803  
 C>Date: 23-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 29-Sep-1999  
 C:Accession: S75967  
 R:Kanevo, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizuo, E.; Nakamura, Y.;  
 Miyajima, N.; Hiroseawa, M.; Sugitara, M.; Sasamoto, S.; Kimura, T.; Hosouchi,



T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpō, S.; Takeuchi, C.; Wada, T.; Matanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.  
 DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.  
 A:Reference number: S74322; MUID:97061201  
 A:Accession: S75967  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-271 <KAN>  
 A:Cross-references: EMBL:D64006; GB:AB001339; NID:g1001291; PIDN:BA010814.1; PID:d1011465; PID:g1001327  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 C:Genetics:  
 A:Start codon: GTG  
 C:Superfamily: conserved hypothetical protein H1191  
 S75967 Length: 271 February 11, 2000 15:51 Type: P Check: 3587 ..

1 MEKCDKNFP GGVREOTKTL PTGCPGSSND HNRLTAESM TKTAIVLLSG  
 51 GDSATVAAL AKREGVRIA LSENYGQRH RELRAADIV QALGIDPHES  
 101 INDLAOMWG SSLTDROCTL PQTGVEPDII PSTYVGRNT VFALGSLA  
 151 EAKAEAVFL GINADISGY PDCRPEYIAT YQOLAASRK VYVGRPIOL  
 201 LAPLEISKV DIYKALELG VPINOTWCSY AGSEPCGRG DSCRLEQAL  
 251 IEAGPELAS AKGRIMREKV D

!!A:SEQUENCE 1.0  
 P1:D69014 - conserved hypothetical protein MTH1108 - *Methanobacterium thermoautotrophicum* (strain Delta H)  
 C:Species: *Methanobacterium thermoautotrophicum*  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 29-Sep-1999  
 C:Accession: D69014  
 R:Smith, D.R.; Doucette-Stamm, L.A.; Delonghery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lum, W.; Pothier, B.; Qiu, D.; Spadefora, R.; Vicare, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso, G.; Bush, D.; Saefer, H.; Patwell, D.; Prabhakar, S.; McDougall, S.; Shimer, G.; Goyal, A.; Pietrowski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
 J. Bacteriol. 179, 7135-7155, 1997  
 A:Title: Complete genome sequence of *Methanobacterium thermoautotrophicum* Delta H: functional analysis and comparative genomics.  
 A:Reference number: A69000; MUID:98037514  
 A:Accession: D69014  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-222 <MTH>  
 A:Cross-references: GB:AE000881; GB:AE000666; NID:g2622206; PIDN:AA85597.1; PID:g2622210  
 A:Experimental source: strain Delta H  
 C:Genetics:  
 A:Gene: MTH1108  
 C:Superfamily: conserved hypothetical protein H1191  
 D69014 Length: 222 February 11, 2000 15:51 Type: P Check: 6260 ..

1 MRAIISISGG MDSAVATAM MDEYEHAIH FDYGRSRM ELEYARLSL  
 51 HLGIEHTLD LOMGLRIGS VLRAGSIPS PSNDDTYEC LERARKWVP  
 101 GRNIVFTSIG VSPAEAMDAG AVYGMWLEE ATEFPDNEE FLDAFNRLLE  
 151 IGLDGVRYV APYIGTKRE IVEAGHEVGL PELLISCYA GDRHVCYCE  
 201 SCMRRRRAFE LAGIDPTEY RE

!!A:SEQUENCE 1.0  
 P1:C65012 - hypothetical protein b2382 - *Escherichia coli* (strain K-12)  
 C:Species: *Escherichia coli*  
 C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 29-Sep-1999  
 C:Accession: C65012  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of *Escherichia coli* K-12.  
 A:Reference number: A64720; MUID:97426617  
 A:Accession: C65012  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-285 <BLAT>  
 A:Cross-references: GB:AE000326; GB:U00096; NID:g1788718; PIDN:AMC75441.1; PID:g1788725; UMG:U00096  
 A:Experimental source: strain K-12, substrain MG1655  
 C:Superfamily: hypothetical protein b2382  
 C65012 Length: 285 February 11, 2000 15:51 Type: P Check: 4986 ..

1 MKAPGLPADQ OFFADLFSG VLNPPQLGRV WFASQASLP VQSLCIDPR  
 51 LDIVLGEVG NILEAKOQL VEGEMLEIPA RAANIPVNNK PYMLSLVFA  
 101 PTWLGSEFY SRTSLHPA ROIQLPSIOR GEGEMALTAI TLHSRPLEQ  
 151 NIIOPLVLSL LHLGRSVNM PCNSOPRGD FLYHSICMW QDNVAPLIR  
 201 ESWAGFNIN PHLKSLQD HGTMREFEY RAVRAKARM ILOKYLHSH  
 251 EVAGRCGFPD SDYFCVFRF QFGLTPEYS ARFG

!!A:SEQUENCE 1.0  
 P1:A70606 - probable electron transfer protein RV3571 - *Mycobacterium tuberculosis* (strain H37Rv)  
 C:Species: *Mycobacterium tuberculosis*  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 29-Sep-1999  
 C:Accession: A70606  
 R:Cole, S.T.; Brosch, P.; Parhail, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigelneter, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Natu, S.; Kiroh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Squares, R.; Sultson, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.  
 A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence.  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: A70606  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-358 <CO>  
 A:Cross-references: GB:292774; GB:AL123456; NID:g3261729; PIDN:CA807145.1; PID:6106711; PID:g1877302  
 A:Experimental source: strain H37Rv  
 C:Genetics:  
 A:Gene: RV3571  
 C:Superfamily: phthalate dioxygenase reductase; cytochrome-b5 reductase  
 A:Keywords: ferredoxin [2Fe-2S] homology  
 F:19-247/Domain: cytochrome-b5 reductase homology <CBR>  
 F:280-344/Domain: ferredoxin [2Fe-2S] homology <CBR>  
 F:305-310,313,343/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted  
 A70606 Length: 358 February 11, 2000 15:51 Type: P Check: 144 ..

1 MTEALGDEPL GDVLEIOIA EVDVDEAR SLVRAVDGS DDEIPRRL

```

51  KRAPQGFLL RYPERGSGV ARCYSLCSPS YTDALAVTV KRTADGYSN
101  WLCDHQVGM RIHVLAPSGN FVPTLDADE LLLAGSGIT PIMSICKSAL
151  AEGGQVTLT YANRDRSVI FGDALRELA KYPDLTVLH WLSLQGLPS
201  ASALAKLVAV YTDPRVEICG PGPFQOARD ALAALKVPAQ QVHIEFKSL
251  ESDPFAAVKV DSDGEAPAT AVVELDGQTH TVSPRTAKL LDVLLAAGLD
301  AFPSCREGHC GACACTLRAG KVMGVNDVL EQDLDESLI LAQSPESD
351  SEVETVDE

!!AA_SEQUENCE 1.0
P1:C65112 - sigma cross-reacting protein 27A - Escherichia coli
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 29-Sep-1999
C:Accession: C65112; JN0287
R:Baltner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;
J.: Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: C65112
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-220 <LAW>
A:Cross-references: GSA:AE000400; GB:U00096; NID:92367203; PIDN:AACT6241.1;
PID:01789602; WMG:103209
A:Experimental source: strain K-12, substrain MG1655
R:Ishihama, R.; Fujita, N.; Ishihama, A.
Biochem. Biophys. Res. Commun. 184, 634-639, 1992
A:Title: Identification of Escherichia coli proteins cross-reacting with
antibodies against region 2.2 peptide of RNA polymerase sigma subunit.
A:Reference number: JN0286; MUID:92246944
A:Accession: JN0287
A:Molecule type: Protein
A:Residues: 4-13, 'X', 15-26, 'I' <UES>
A:Comment: This protein reacts with antiserum against a synthetic peptide of
RNA polymerase sigma chains.
C:Genetics:
A:Gene: yhbI
C:Superfamily: sigma cross-reacting protein 27A

C65112 Length: 220 February 11, 2000 15:51 Type: P Check: 597 ..

1  MITMKRIGV LSGGYDGS EHEAVLTL AISRGAQV CFAPDKQVD
51  VINHLTGEM TETRNVLIEA ARTREIRP LAQDAEALD ALIVPGFGA
101  AKNLNFSASL GSECTVREL KALAQMHOA GKPLGFMCA PAMLPRIIDF
151  PLRLITGTDI DTAEVLEMG AEHVPCPVD IVDENKIV TTPAYMLAQN
201  IAEAASGIDK LVSRVYLAE

!!AA_SEQUENCE 1.0
P1:JC5603 - esi protein - zebra fish
C:Species: Brachydanio rerio (zebra fish)
C>Date: 23-Sep-1997 #sequence_revision 23-Sep-1997 #text_change 29-Sep-1999
C:Accession: JC5603; PC4488
R:Chang, H.; Gilbert, W.
Biochem. Biophys. Res. Commun. 237, 84-89, 1997
A:Title: A novel zebrafish gene expressed specifically in the photoreceptor
cells of the retina.
A:Reference number: JC5603; MUID:97410119
A:Accession: JC5603
A:Molecule type: mRNA
A:Residues: 1-270 <CHA>
A:Cross-references: GB:U10403; NID:9499129; PIDN:AA60261.1; PID:9499130
A:Experimental source: eye

```

```

A:Accession: PC4488
A:Molecule type: protein
A:Residues: 105-137 <CH2>
C:Comment: This protein is important for maintaining normal retina structure
and function.
C:Genetics:
A:Gene: esi
C:Superfamily: sigma cross-reacting protein 27A

JC5603 Length: 270 February 11, 2000 15:51 Type: P Check: 1627 ..

1  MLASRALIAK QAAAMLVROP ACLMHGGDW GNNGNINIAV VSGCGWMDG
51  TDIHEAAVTM YHLSRNGARF QIFAPNOQM HYMDHAKMP SSDNRNIM
101  ESARFSGOG MQQNDLSKL DANSPDAVF PGCHGIYKNN STESKDGDC
151  KLNDDVERVL KDFHARKPI GLSSMAPLA CRVLPSEVT MGERDESSR
201  WGRPNNTMAY QAVKSGARH NTRPEYATV DEKNVISTP TFMWETDVHY
251  HYFDGIGNM VKHVMKTK

!!AA_SEQUENCE 1.0
P1:T09924 - cytidine deaminase (EC 3.5.4.5) CDA6 - Arabidopsis thaliana
N:Alternate names: Protein T16L4.120
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Sep-1999
C:Accession: T09924
R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Bancroft, I.; Mewes, H.W.;
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z16897
A:Accession: T09924
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-293 <BEV>
A:Cross-references: EMBL:AL079344; GSPDB:GN00062; ATSP:T16L4.120
A:Experimental source: cultivar Columbia; BAC clone T16L4
C:Genetics:
A:Gene: ATSP:T16L4.120
A:Map position: 4
C:Superfamily: cdd protein
C:Keywords: hydrolase

T09924 Length: 293 February 11, 2000 15:51 Type: P Check: 1428 ..

1  MKFVTPSEA AEGSVRPSD LPKLIDKAMS LARAPVTFK VQAVGLTSSG
51  EYFLGVNVEF PNLPLHHTH AEQFLVTLA LNSKKLITHI AVSVGTITFG
101  APCGHCQFV QEMRNAPETE ILIKRPKDI DEFNSLSLM PRFPDSIL
151  PEDASLLEQ RDSNVLSDP EICSDPEDC SHTKRALAA ANKSVAPYSK
201  CPSEVALICG GEYKGYITE SVAINPSLGP VEALYDEVA RCGKREFNEI
251  TEYVLVEKMD VAVSQEATAR TFLDKIAPKC DFKVLCHYKT NKN

!!AA_SEQUENCE 1.0
P1:S5551 - cytokine-inducible protein CIS - mouse
C:Species: Mus musculus (house mouse)
C>Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 28-May-1999
C:Accession: S5551
R:Yoshimura, A.; Ohnubo, T.; Kiyuchi, T.; Jenkins, N.A.; Gilbert, D.J.;
Copeland, N.G.; Hara, T.; Miyajima, A.
EMBO J. 14, 2816-2826, 1995
A:Title: A novel cytokine-inducible gene CIS encodes an SH2-containing protein
that binds to tyrosine-phosphorylated interleukin 3 and erythropoietin
receptors.
A:Reference number: S5551; MUID:95317300
A:Accession: S5551
A:Status: preliminary

```

A:Molecule type: mRNA  
 A:Residues: 1-257 <YOS>  
 A:Cross-references: GB:D31943; NID:q1041128; PIDN:BA06713.1; PID:dl007285;  
 PID:q1041129  
 C:Superfamily: cytokine-inducible protein C1S; SH2 homology  
 F:82-178/Domain: SH2 homology <SH2>  
 S55551 Length: 257 February 11, 2000 15:51 Type: P Check: 2728 ..

1 MYLCVQSCSP LLAVEIGRR PLMAQSLPL GPAMQPLPTG AFPEVTEET  
 51 PVQAKENPKV LDPEGLDLCI AKTSYLRES GWWSGITSAS EARGHLOKMP  
 101 EGTLYVDST HPSYLFITLV KTRGPNVR IEVADSFL DSNLSRPRI  
 151 LAFPDVSLV QHYVASCAD TRSDSPDAP TPALPMSKOD APSDSVLPID  
 201 VATAVHLKV QPVRSSNR SLQHLRLVT NRIYAVDCL PLPRMDYL  
 251 RQPPQL

!!AA\_SEQUENCE 1.0  
 P1:JC5626 - STAT induced STAT inhibitor 2 - human  
 C:Species: Homo sapiens (man)  
 C>Date: 14-Oct-1997 #sequence\_revision 07-Nov-1997 #text\_change 07-May-1999  
 C:Accession: JC5626  
 R:Matsumoto, S.; Ikegami, K.; Ueno, K.; Narazaki, M.; Naka, T.; Yamamoto, H.;  
 Matsumoto, T.; Saito, H.; Hosoe, S.; Kishimoto, T.  
 Biochem. Biophys. Res. Commun. 237, 79-83, 1997  
 A:Title: Cloning and functional analysis of new members of STAT induced STAT  
 inhibitor (SSI) family: SSI-2 and SSI-3  
 A:Reference number: JC5626; MUID:97410118  
 A:Accession: JC5626  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-198 <MIN>  
 C:Comment: This protein plays a role in negative feedback control of Janus  
 kinase-signal transduction and activator of transcription signaling pathway.  
 C:Superfamily: cytokine-inducible protein C1S; SH2 homology  
 F:48-149/Domain: SH2 homology <SH2>  
 F:163-172/Region: SC motif 1  
 F:182-192/Region: SC motif 2  
 JC5626 Length: 198 February 11, 2000 15:51 Type: P Check: 1738 ..

1 MTLRCLPESG NGEGTRSQW GTAGSAEPPS PQARLAKAL RELQGTGWTW  
 51 GSMIVNEAKE KLEAPEGTF LIRDSHSDY LLTISVYTS A GPTNIRIYQ  
 101 DGRFLDSII CVKSKLKQFD SVVHLIDYV QMKCKRTDP EAPRNGIVHL  
 151 YLTKPLYTSA PSLQHLCLRL INKCTGAIWG LPLPRLKDY LEEYKFOV

!!AA\_SEQUENCE 1.0  
 P1:JC5760 - cytokine-inducible SH2 protein 2 - human  
 C:Species: Homo sapiens (man)  
 C>Date: 24-Jan-1998 #sequence\_revision 13-Mar-1998 #text\_change 07-May-1999  
 C:Accession: JC5760  
 R:Masuhara, M.; Sakamoto, H.; Matsumoto, A.; Suzuki, R.; Yasukawa, H.; Mitsui,  
 K.; Wakioka, T.; Tanimura, S.; Sasaki, A.; Mitsuwa, H.; Yokouchi, M.; Ohtsubo,  
 M.; Yoshimura, A.  
 Biochem. Biophys. Res. Commun. 239, 439-446, 1997  
 A:Title: Cloning and characterization of novel C1S family genes.  
 A:Reference number: JC5760; MUID:98008857  
 A:Accession: JC5760  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-198 <MAS>  
 A:Cross-references: DDBJ:AB006966  
 C:Comment: This protein plays a role in the negative regulation of cytokine  
 signaling by interacting with specific targets.  
 C:Superfamily: cytokine-inducible protein C1S; SH2 homology  
 F:47-142/Domain: SH2 status predicted <SH2>

JC5760 Length: 198 February 11, 2000 15:51 Type: P Check: 1381 ..

1 MTLRCLPESG NGEGTRSQW GTAGSAEPPS PQARLAKAL RELQGTGWTW  
 51 GSMIVNEAKE KLEAPEGTF LIRDSHSDY LLTISVYTS A GPTNIRIYQ  
 101 DGRFLDSII CVKSKLKQFD SVVHLIDYV QMKCKRTDP EAPRNGIVHL  
 151 YLTKPLYTSA PSLQHLCLRL INKCTGAIWG LPLPRLKDY LEEYKFOV

!!AA\_SEQUENCE 1.0  
 P1:A41900 - cyn operon regulatory protein cynR - Escherichia coli  
 C:Species: Escherichia coli  
 C>Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 29-Sep-1999  
 C:Accession: A41900; B64761  
 R:Sung, Y.C.; Fuchs, J.A.  
 J. Bacteriol. 174, 3645-3650, 1992  
 A:Title: The Escherichia coli K-12 cyn operon is positively regulated by a  
 member of the IYR family.  
 A:Reference number: A41900; MUID:92276346  
 A:Contents: K-12  
 A:Accession: A41900  
 A:Molecule type: DNA  
 A:Residues: 1-739 <SUN>  
 A:Cross-references: GB:W93053; NID:q145645; PIDN:AA23628.1; PID:q145646  
 A:Note: Sequence extracted from NCBI backbone (NCBIN:104623, NCBI:P:104625)  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;  
 Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,  
 J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:97426517  
 A:Accession: B64761  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-299 <BLAT>  
 A:Cross-references: GB:AO00141; GB:U00096; NID:q1786532; PIDN:AC73441.1;  
 PID:q1786533; UMG:P:D0338  
 A:Experimental source: strain K-12, substrain MG1655  
 C:Genetics:  
 A:Gene: cynR  
 C:Superfamily: probable transcription regulator IyR  
 C:Keywords: DNA binding; transcription regulation  
 A41900 Length: 299 February 11, 2000 15:51 Type: P Check: 4304 ..

1 MLSRHINFL AVAHEGSFTR AASALVHQP ALSQOIROLE ESUGVPLFDR  
 51 SGRITLITDA GEVWROYASR ALQELGAGKR AIHDVADLRL GSLRIAVPT  
 101 FTSYFEGPLM ADFYARFPI TLQLOMSQE KIEMLCRDE LDVGIAPAV  
 151 HSPLEFAIPL LTESIALVA QHHPPLAVHQ VALSRINDER LVLSAEFAT  
 201 REQIDHYCKR AGHPQVYIE ANSISAVLEL IRRISLSTL PAIAIHOHG  
 251 LKAISLAPPL LERRAVILRR KNSWQTAANK AFLMALDKC AAYGQNESR

!!AA\_SEQUENCE 1.0  
 P1:F70721 - hypothetical protein RV1829 - Mycobacterium tuberculosis (strain  
 H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 26-Aug-1999  
 C:Accession: F70721  
 R:Coile, S.T.; Brosemer, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.;  
 Gordon, S.V.; Eigmeier, K.; Gass, S.; Barry III, C.E.; Tekala, F.; Badcock, K.;  
 Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, T.; Devlin, K.;  
 Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagers, K.;  
 Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail,  
 M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.;  
 Squares, S.  
 Nature 393, 537-544, 1998

A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: F70721  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-164 <COL>  
 A:Cross-references: GB:278020; GB:AL123456; NID:g3261625; PID:e1300004; PID:g3261627  
 A:Experimental source: strain H37Rv  
 C:Genetics:  
 A:Gene: RV1829  
 C:Superfamily: Synecocystis conserved hypothetical protein s111142  
 F70721 Length: 164 February 11, 2000 15:51 Type: P Check: 7074 ..

1 MGEVAVGIR VEQPQNPVL LREANGDRY LPWIGOSEA AA1ALEQGV  
 51 EPPRPETHDL IRDLIALGH SLKEVRIYDL QEGTFVADLI FDRNRYVSAR  
 101 PSDVAIALR VGVPIYEEA VLAQAGLLIP DESDEATTA VREDEVEKER  
 151 EFLDSVSPDD FKAT

!!AA\_SEQUENCE 1.0  
 P1:S56618 - yjx protein - Escherichia coli  
 C:Species: Escherichia coli  
 C>Date: 28-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 18-Sep-1998  
 A:Accession: S56618; A65255  
 R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.  
 Nucleic Acids Res. 23, 2105-2119, 1995  
 A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.8 through 100 minutes.  
 A:Reference number: S56314; MUID:95334362  
 A:Accession: S56618  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-173 <BUR>  
 A:Cross-references: EMBL:U14003; NID:g1263172; PID:g537234  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:97426617  
 A:Accession: A65255  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-173 <BLAT>  
 A:Cross-references: GB:AEO00509; GB:U00096; NID:92367383; PID:g1790855; MGI:104394  
 A:Experimental source: strain K-12, substrain MG1655  
 C:Genetics:  
 A:Gene: yjx  
 C:Superfamily: Escherichia coli conserved yjx protein  
 S56618 Length: 173 February 11, 2000 15:51 Type: P Check: 3873 ..

1 MLIMQVCA TTNPAKIOAI LQAFHEIFGE GSCHIASVAV ESGVDEPFG  
 51 SEETRAGARN RVANARRLLP EADFVAIEA GIDGOSTSW VVIMASORG  
 101 EARSATLPLP AVILEKVBEG EALGPVMSRY TGIDIGRKE GAIGVFTAGK  
 151 LTRASVYHOA VIALSPFN AVY

C>Date: 07-May-1998 #sequence\_revision 15-May-1998 #text\_change 18-Sep-1998  
 C:Accession: S45255  
 R:Arvidson, D.N.; Arvidson, C.G.; Lawson, C.L.; Miner, J.; Adams, C.; Underhill, P.  
 Nucleic Acids Res. 22, 1821-1829, 1994  
 A:Title: The tryptophan repressor sequence is highly conserved among the Enterobacteriaceae.  
 A:Reference number: S45254; MUID:94268903  
 A:Accession: S45255  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-54 <ARV>  
 A:Cross-references: EMBL:L26582; NID:g433053; PID:g433055  
 A:Experimental source: ATCC 13048  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993  
 C:Superfamily: Escherichia coli conserved yjx protein  
 S45255 Length: 54 February 11, 2000 15:51 Type: P Check: 4167 ..

1 EALGPVMSOH TGIDIGRKE GAIGVFTAGK LTRASVYHOA VIALSPFN  
 51 AIYR

!!AA\_SEQUENCE 1.0  
 P1:A70738 - probable rlmI protein - Mycobacterium tuberculosis (strain H37Rv)  
 C:Species: Mycobacterium tuberculosis  
 C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 24-Sep-1999  
 A:Accession: A70738  
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Bauman, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feldwell, I.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Hornsby, T.; Jagels, K.; Krogan, A.; McLellan, J.; Molle, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: A70738  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-158 <COL>  
 A:Cross-references: GB:Z77165; GB:AL123456; NID:g3261609; PID:e255179; PID:g1449367  
 A:Experimental source: strain H37Rv  
 C:Genetics:  
 A:Gene: rlmI  
 C:Superfamily: Escherichia coli ribosomal-protein-alanine N-acetyltransferase rlmI  
 A70738 Length: 158 February 11, 2000 15:51 Type: P Check: 8373 ..

1 MTDITPEVPI GALTIRADQR CAELERQLEY GDDPMPAAF NRELASPHNH  
 51 YVARSBGTL VGVAGISRLG RPPPEVEVA TIGVPAYOG RGIGRRLRE  
 101 LLDFAAGVY YLEVRTDND AALALYSVGF QVGLRRRY RVSGADAYTM  
 151 RRDGDPDS

!!AA\_SEQUENCE 1.0  
 P1:F72316 - hypothetical protein TM0917 - Thermotoga maritima (strain MS8)  
 C:Species: Thermotoga maritima  
 C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 11-Jun-1999  
 A:Accession: F72316  
 R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Nelson, W.C.; Ketchum, K.A.; McDonald, L.; Utterback, T.R.; Malek, J.A.; Linher, K.D.; Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; Heidelberg, J.; Sutton, G.G.; Fleischmann, R.D.; White, O.; Salzberg, S.L.; Smith, H.O.;

Venter, J.C.; Fraser, C.M.  
 Nature 399, 333-329, 1999  
 A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from  
 genome sequence of *Thermococcus maritima*.  
 A:Reference number: A72200; MUID:99287316  
 A:Accession: F72316  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-309 <ARN>  
 A:Cross-references: GB:AE001756; GB:AE000512; NID:94981453; PID:94981454;  
 A:Experimental source: strain MSB8  
 A:Genetics:  
 C:Gene: TM0917  
 F72316 Length: 309 February 11, 2000 15:51 Type: P Check: 5138 ..

1 MFWLLPSLF LGWSLGANDA ANVEGPEVGS GLIPYKRAI VASIFVLGS  
 51 VLGARGLQI ISSLSISDL LSIIVLSGA LVTIMTKLG IPVTSQAV  
 101 GGIIQANTV MGIGIDESA LTKILTWFL TPVGAPFLS IPYPLSLF  
 151 RKIPSIQID RVIKISAMII GAYGAFSLGA NNVAANTGVF AKKILITCA  
 201 ARLGISIAI GILYSKNM LTVGNLIEL DFTSLIAVL SQAMVWIFS  
 251 LIGIHSSSO AIVGAVLAG YSKGMNLGN KVLKILSGV FLTPAVSGTL  
 301 SFLTSLIK

11AA:SEQUENCE 1.0  
 P1:B70451 - prephenate dehydrogenase - *Aquifex aeolicus*  
 C:Species: *Aquifex aeolicus*  
 C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 08-May-1998  
 C:Accession: B70451  
 R:Beckert, G.; Warren, P.V.; Gaasterland, T.; Young, M.G.; Lenox, A.L.; Graham,  
 D.E.; Overbeek, R.; Sneed, M.A.; Keller, M.; Anjey, M.; Huber, R.; Feldman,  
 R.A.; Short, J.M.; Olson, G.J.; Swanson, R.V.  
 Nature 392, 353-358, 1998  
 A:Title: The complete genome of the hyperthermophilic bacterium *Aquifex*  
*aeolicus*.  
 A:Reference number: A70300; MUID:98196666  
 A:Accession: B70451  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-311 <AOE>  
 A:Cross-references: GB:AE000754; NID:92984047; PID:92984048; GB:AE000657  
 A:Experimental source: strain VFS  
 C:Genetics:  
 A:Gene: tyra  
 B70451 Length: 311 February 11, 2000 15:51 Type: P Check: 4885 ..

1 MALISMENP SPPOGCRKN IIKLKSLSM QNVILVYGF MGGSFAKSLR  
 51 RSGGKRIY YDINPESISK AVDLGIIDEG TSIKAVEDE SPDEFMLSP  
 101 VRTREIRAK LSYLSEDAI VTDOGSVKG LVDLENLIG KREYGHPIA  
 151 GTERSGEYS LDNIYEGKV ILPTKTKDK KRLKIVRWY EDVGVEYEM  
 201 SPELDHYFG VVSHLPYAVA FALVDLILM STPEVDLFKY PGGGKDFTR  
 251 IAKSPTIMR DIFLENKRV MKAIEGEKS LNLKELIYR EAEELVEYL  
 301 KEVIRKMEI D

11AA:SEQUENCE 1.0  
 P1:G39845 - dihydroacetate dehydrogenase (electron transfer subunit) *pyrDII* -  
*Bacillus subtilis*  
 C:Species: *Bacillus subtilis*  
 C:Date: 20-Mar-1992 #sequence\_revision 10-Apr-1992 #text\_change 24-Sep-1998

C:Accession: G39845; D69686  
 R:Quinn, C.L.; Stephenson, B.T.; Switzer, R.L.  
 J Biol Chem 265, 9133-9127, 1991  
 A:Title: Functional organization and nucleotide sequence of the *Bacillus*  
*subtilis* pyrimidine biosynthetic operon.  
 A:Reference number: A39845; MUID:91225016  
 A:Accession: G39845  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-256 <OUT>  
 A:Cross-references: GB:M59757; NID:9387576; PID:9143391  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo,  
 V.; Bertero, M.G.; Besieres, P.; Bilotin, A.; Borchert, S.; Borriest, R.;  
 Boursier, L.; Brans, A.; Braun, M.; Brignelli, S.C.; Broo, S.; Brouillet, S.;  
 Burschl, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani,  
 J.J.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.;  
 Dueterhoeft, A.; Ehrlich, S.D.; Emmerson, P.T.; Eutian, K.D.; Errington, J.;  
 Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallazzi,  
 A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Gollightly, E.J.; Grandi,  
 G.; Giuseppe, G.; Guy, B.J.; Haga, K.; Haeche, J.; Hatwood, G.R.; Henaut, A.;  
 Hilbert, H.; Holappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.;  
 Joris, B.; Karamata, D.; Kasahara, Y.; Klaser-Biancard, M.; Klein, C.; Kurita, K.;  
 Kobayashi, Y.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.;  
 Lapidus, A.; Lardinols, S.  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda,  
 S.; Mausel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.;  
 Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogawa, A.; Oudega,  
 B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Portwolk, S.; Prescott,  
 A.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds,  
 T.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato,  
 T.; Scanlon, P.  
 A:Authors: Schleich, S.; Schroter, R.; Scofield, F.; Sekiguchi, J.; Sekowska,  
 A.; Serró, S.J.; Serró, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Taccini, E.;  
 Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka,  
 T.; Terpe, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenoel, M.;  
 Vanlier, F.; Vassartelli, A.; Viani, A.; Wamburt, R.; Wedler, E.; Wedler, H.;  
 Wetzenecker, T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto,  
 K.; Yata, K.; Yoshida, K.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Authors: Yoshikawa, H.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus*  
*subtilis*.  
 A:Reference number: A69580; MUID:98044033  
 A:Accession: D69686  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-256 <KUN>  
 A:Cross-references: GB:Z99112; GB:AL009126; NID:92633902; PID:9185145;  
 PID:92633926  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: pyrDII  
 G39845 Length: 256 February 11, 2000 15:51 Type: P Check: 2638 ..

1 MKRAYLVCS NOCIADRVFO MVKLGELYOG FTTPOGFLHL KYSEAVTPLL  
 51 RRPISADVN FEKNVITILY RVDSGETRLI SLKQOGLYD VLGPLNGFPP  
 101 VNEVQPKTA LLYGGVGVV PLOELSKRLI EKVAVIIVHL GFOSAKDVFX  
 151 EEECRQYGDY YVATADGSYG ETGFVDIVIK RKLLEFDILL SCGTPMLKA  
 201 LKQYAKREV YLSMERMGCG GIGACFACVC HNESETSYV KYCLDGVPFK  
 251 AQEVAL

11AA:SEQUENCE 1.0  
 P1:J01382 - hypothetical 34k protein - pea enation mosaic virus  
 C:Species: pea enation mosaic virus; PEKV  
 C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 26-Aug-1999  
 C:Accession: J01382

R:Demler, S.A.; de Zoeten, G.A.  
 J. Gen. Virol. 72, 1819-1834, 1991  
 A:Title: The nucleotide sequence and luteovirus-like nature of RNA 1 of an aphid non-transmissible strain of pea enation mosaic virus.  
 C:Reference number: J01382; MUID:91341468  
 A:Accession: J01382  
 A:Molecule type: genomic RNA  
 A:Residues: 1-303 <DEM>  
 A:Cross-references: GB:104573; NID:9294105; PID:9294108  
 A:Experimental source: strain MSG  
 A:Note: 254-His was also found  
 C:Genetics:  
 A:Map position: segment RNA1  
 C:Superfamily: pea enation mosaic virus hypothetical 34k protein  
 J01382 Length: 303 February 11, 2000 15:51 Type: P Check: 6785

1 MHGIDQPLP LDYVRCASL SFLASLDGL LSEARELSGP LALITSSYLL  
 51 LVSLALCKAI PGSTWYRPG WLQPVSGRWL IFCGPTSLAQ RFRLYAARLG  
 101 LVLSNCRPH GQSAITLQS YNALPNNIWM DMAOLDILTF SMPIANTFAY  
 151 LADCARPP IVEGVSAY VPTLGLTHQ DPRLYALRR RNLDLGEPR  
 201 RVRQVEESM ALLCSSVRSST SRSROIPLY GSVLHVYGL AERDCILFPT  
 251 DSNSSSYTHR VLEODRNRAD QSLFSIDLEY VHDELIALG YSDEDEDLD  
 301 NFF

11AA\_SEQUENCE 1.0  
 P1:S56961 - Probable membrane protein YUL178c - Yeast (Saccharomyces cerevisiae)  
 N:Alternate names: hypothetical protein J0490  
 C:Species: Saccharomyces cerevisiae  
 C>Date: 08-Jul-1995 #sequence\_revision 08-Sep-1995 #text\_change 26-Aug-1999  
 C:Accession: S56961  
 R:Obermaier, B.; Piravandi, E.; Rinke, M.; Domdey, H.  
 submitted to the Protein Sequence Database, September 1995  
 A:Reference number: S56937  
 A:Accession: S56961  
 A:Molecule type: DNA  
 A:Residues: 1-196 <OBE>  
 A:Cross-references: EMBL:249453; NID:q1008377; PID:q1008378; MIPS:YUL178C  
 C:Genetics:  
 A:Map position: 10L  
 C:Superfamily: Saccharomyces probable membrane protein YUL178c  
 C:Keywords: transmembrane protein  
 S56961 Length: 196 February 11, 2000 15:51 Type: P Check: 8234

1 MCGGLDVL PGKAITTQI IDPKNTGN VEETESALTL TLKATWGAN  
 51 SFAKLEFQC NDNMKODELT SHTWAKSIQ LTLKSGSGL KSKDDKNG  
 101 DGNKGKDGDS ESKPAKAG GTSNFWLFL YALLFLIYL MYVEFLNTRG  
 151 GSVQDRAEF IQRSTOFLTS LPEFCREYVS RILGSAOR GGYSAV

11AA\_SEQUENCE 1.0  
 P1:J01355 - V0 protein - Miscanthus streak virus  
 C:Species: Miscanthus streak virus  
 C>Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 20-Sep-1999  
 C:Accession: J01355; J00918  
 R:Chatant, M.; Matsumoto, Y.; Mizuta, H.; Ikegami, M.; Boulton, M.I.; Davies, J.W.  
 J. Gen. Virol. 72, 2325-2331, 1991  
 A:Title: The nucleotide sequence and genome structure of the geminivirus miscanthus streak virus.  
 C:Reference number: J01355; MUID:92013947  
 A:Accession: J01355  
 A:Molecule type: DNA  
 A:Residues: 1-134 <CHA>

A:Cross-references: DDBJ:D01030; NID:9222128; PID:01001300; PID:9222129  
 C:Comment: Miscanthus streak virus causes leaf streak.  
 C:Superfamily: Miscanthus streak virus V0 protein  
 J01355 Length: 134 February 11, 2000 15:51 Type: P Check: 9413

1 MCGVIMGR GSENNITAP PWPISRPDA SVRCAGPRS LPLADRDQ  
 51 AAVKWRHLS VFALVPLVC FKRRRLRDR TKGTQVAC LRTWVAKFCF  
 101 VCKFLIPMV AVAILMNVH CMVYLWHLR CFAS

11AA\_SEQUENCE 1.0  
 P1:J00150 - hypothetical 13k protein - Pseudomonas aeruginosa  
 C:Species: Pseudomonas aeruginosa  
 C>Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 20-Sep-1999  
 C:Accession: J00150  
 R:Kato, J.; Chu, L.; Kitano, K.; Devault, J.D.; Kimbara, K.; Chakrabarty, A.M.; Misra, T.K.  
 Gene 84, 31-38, 1989  
 A:Title: Nucleotide sequence of a regulatory region controlling alginate synthesis in Pseudomonas aeruginosa characterization of the algR2 gene.  
 A:Reference number: J00132; MUID:90108714  
 A:Accession: J00150  
 A:Status: translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-122 <KAT>  
 C:Superfamily: Pseudomonas aeruginosa hypothetical 13k protein  
 J00150 Length: 122 February 11, 2000 15:51 Type: P Check: 6657

1 MALSPACSR CCAASARP ATACCPGA SRTCARSP WIAMWIRCS  
 51 TATVACARS RMPPPKRT TLPWRACIP NMTPTVTPP MHGRSCWPD  
 101 SASASRWSA GSVSPAVG CA

11AA\_SEQUENCE 1.0  
 P1:B71217 - hypothetical protein PH2002 - Pyrococcus horikoshii  
 C:Species: Pyrococcus horikoshii  
 C>Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 20-Sep-1999  
 C:Accession: B71217  
 R:Kawabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kotsugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohnuki, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kishida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.  
 DNA Res. 5, 55-76, 1998  
 A:Title: Complete sequence and gene organization of the genome of a hyperthermophilic archaebacterium, Pyrococcus horikoshii OT3.  
 A:Reference number: A71000; MUID:96344137  
 A:Accession: B71217  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-108 <KAN>  
 A:Cross-references: GB:AP000007; GB:AP000001; NID:93236134; NID:93236128; PID:01032072; PID:01030009; PID:93258446; PID:93263683  
 A:Experimental source: strain OT3  
 A:Note: This accession replaces an interim accession for a sequence replaced by Genbank  
 A:Note: This sequence is split into two separate translations in Genbank (residues 1-37, 38-108)  
 C:Genetics:  
 A:Gene: PH2002  
 C:Superfamily: Pyrococcus horikoshii hypothetical protein PH2002  
 B71217 Length: 108 February 11, 2000 15:51 Type: P Check: 8463

1 MGKHYLEKS ILRLKIDIF KVCHLDPLC EYTLGSGFS LHRFDLFP  
 51 VQGHRRHPSL VFILKIQGV KEDIFKLLT PSRSQAVRH KPPHPPFVQ

## 101 SSDPSICP

11AA:SEQUENCE 1.0  
 PI:T05820 - hypothetical protein TSK18.150 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 20-Sep-1999  
 C:Accession: T05820  
 R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voel, M.; Robben, J.;  
 Volckaert, G.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schellier, C.  
 submitted to the Protein Sequence Database, April 1998  
 A:Reference number: 215453  
 A:Accession: T05820  
 A:Molecule type: DNA  
 A:Residues: 1-217 <BEV>  
 A:Cross-references: EMBL:AL022560  
 A:Experimental source: cultured Columbia; BAC clone TSK18  
 C:Genetics:  
 A:Map position: 4  
 A:Insertions: 30/3; 101/2  
 A:Note: TSK18.150  
 C:Superfamily: Arabidopsis thaliana hypothetical protein TSK18.150

T05820 Length: 217 February 11, 2000 15:51 Type: P Check: 4391

- 1 MNHFLYSV FSGLVSEFT CFAAEFKRTQ KEDIRMDTER NCYVPGSHAF
- 51 GLGSAAVLCF CLAQIVGNIV VERNHRTRTK REDGYKITDL TLPTVLLLS
- 101 MSNFVVVLLI LSTAFMSRA QAYEGWIDE DCYLVDGVF AASGLAIIIG
- 151 LGALTIATNR IKVKKQOOLV QVYIKDONQD QRRSMEEOK HDEHQINKSE
- 201 SVIHVEVS STNISRI

11AA:SEQUENCE 1.0  
 PI:S77462 - hypothetical protein s11071 - Synechocystis sp. (strain PCC 6803)  
 C:Species: Synechocystis sp.  
 A:Variety: PCC 6803  
 C>Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Sep-1999  
 C:Accession: S77462  
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.;  
 Miyajima, N.; Hirose, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi,  
 T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shampo, S.;  
 Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.  
 DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium  
 Synechocystis sp. PCC6803. II. Sequence determination of the entire genome and  
 assignment of potential protein-coding regions.  
 A:Reference number: S74322; MUID:97061201  
 A:Accession: S77462  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-275 <KAN>  
 A:Cross-references: EMBL:D90905; GB:AB001339; NID:91652360; PID:d1018042;  
 PID:91652387  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June  
 1996  
 C:Superfamily: Synechocystis hypothetical protein s11071

S77462 Length: 275 February 11, 2000 15:51 Type: P Check: 3256

- 1 MAHLPSSAT RSKNFVWL AASLSLMLM GPASPAQAVN NPELFRKRT
- 51 PVVDLAFLP EIOEALIDD LNSPEVETGW KLRVLTQYDR SPERAVIPEW
- 101 GLDDKSLIV ADARGNLLA FSIODEVYEL MPTWIMEMQ ARFGNMYYIR
- 151 DNGENLAITE ALETYKGLL KGCNNVPGI PREOMILIVY TSIYGLIFG
- 201 FAALPSPNO TFMQWVLM SPLMGILVIA FGIQPVYRTI SDFLEPRNL
- 251 MGFSIGALVA YLSPMFSQIN NNPOI

11AA:SEQUENCE 1.0  
 PI:S53572 - probable membrane protein Y1L071w-a - yeast (Saccharomyces cerevisiae)

C:Species: Saccharomyces cerevisiae  
 C>Date: 06-May-1995 #sequence\_revision 19-Oct-1995 #text\_change 20-Sep-1999  
 C:Accession: S53572  
 R:Smith, V.  
 submitted to the EMBL Data Library, September 1994  
 A:Reference number: S48407  
 A:Accession: S53572  
 A:Molecule type: DNA  
 A:Residues: 1-158 <SM>  
 A:Cross-references: EMBL:Z38060; MIPS:Y1L071w-a  
 C:Genetics:  
 A:Map position: 9L  
 C:Superfamily: Saccharomyces probable membrane protein Y1L071w-a  
 C:Keywords: transmembrane protein  
 F:58-74/Domain: transmembrane #status predicted <TM1>  
 F:102-118/Domain: transmembrane #status predicted <TM2>

S53572 Length: 158 February 11, 2000 15:51 Type: P Check: 4181

- 1 MNKVTLLP RVFCLMSV MWNDRKSD RSVNLDQHS NKPPSESLIS
- 51 LLSLSMSI SLSLLIIV AVLEPPLYS GLPMIMSEKS VSFESLSIT
- 101 SFIVPELII SLTIICKIEL SSKCITPSI SKFNCPDIIN FVNSLSISK
- 151 SIPEVDCK

11AA:SEQUENCE 1.0  
 PI:G45355 - ORF1 protein - Autographa californica nuclear polyhedrosis virus  
 C:Species: Autographa californica nuclear polyhedrosis virus, AcMNPV  
 C>Date: 28-Apr-1993 #sequence\_revision 28-Apr-1993 #text\_change 09-Sep-1997  
 C:Accession: G45355  
 R:Guarino L.A.; Smith, M.W.  
 Virology 175, 1-8, 1990  
 A:Title: Nucleotide sequence and characterization of the 39K gene region of  
 Autographa californica nuclear polyhedrosis virus.  
 A:Reference number: A45355; MUID:91021010  
 A:Accession: G45355  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-146 <GUA>  
 A:Cross-references: GB:M37122; NID:9332383; PID:9332390

G45355 Length: 146 February 11, 2000 15:51 Type: P Check: 1335

- 1 MWTLQDPILY AYAVHDGAKR TKIAAFDIDG TLISSTREK FPKNPDDMQL
- 51 LPCAKRLKRL YELGYDLVVF TNOAHLGSGK IKASDLIYKL ENIKKATGVP
- 101 ISFYVSPNKD EHRKPTRECG AHPNSLRIL TRNNLFMMAT RLAEILI

11AA:SEQUENCE 1.0  
 PI:S32975 - gene BCRF2 protein - human herpesvirus 4  
 N:Alternate names: gene BMRF1 protein  
 C:Species: human herpesvirus 4, Epstein-Barr virus  
 C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 13-Mar-1998  
 C:Accession: S32975; S32976; S32977; S32978; S32980; S32981; S32982;  
 S32983; S32984; S32985; S32986  
 R:Farrell P.J.  
 submitted to the EMBL Data Library, March 1988  
 A:Reference number: S32973  
 A:Accession: S32975  
 A:Molecule type: DNA  
 A:Residues: 1-363 <FAR>  
 A:Cross-references: EMBL:V01555; NID:959074; PID:91334836;  
 PID:925078; PID:91334837; PID:925079; PID:91334838; PID:925080; PID:91334839;  
 PID:925081; PID:91334840; PID:925066; PID:91334841; PID:925067; PID:91334842;  
 PID:925068; PID:91334843; PID:925069; PID:91334844; PID:925070; PID:91334845;  
 PID:925071; PID:91334846; PID:925072; PID:91334847  
 A:Note: each of the twelve author supplied translations in EMBL:V01555 for this

repeated gene is marked as conflicting with the conceptual translation

C:Genetics:  
A:Gene: BCRR2\_1; BMRFL\_2; BMRFL\_3; BMRFL\_4; BMRFL\_5; BMRFL\_6; BMRFL\_7; BMRFL\_8;  
BMRFL\_9; BMRFL\_10; BMRFL\_11; BMRFL\_12  
A:Note: twelve consecutive ORFs apparently encode the identical polypeptide

S32975 Length: 383 February 11, 2000 15:51 Type: P Check: 4975 ..

1 VWEAGRRPR GEVEDRRPGL CMQSPGDPR PSQSPSRSA PQDPRSRQ  
51 GPASSGAAGS PPOAPOTRVS ASRADRRAM RLIGASRRGW FCPSLPSEE  
101 PGTSGTPEPL GPASRRPGL RSLSPVKPK ECLGATLGA QAPESGQGH  
151 LRVPRVPGQ PEGPRCPGRP ORVPRPFG LQSPGCEPG TLGVSPPLQ  
201 ARASSRRGA SLGPOVPHR DSGDPPTG PSLCPAPLQ PSLHRRPQL  
251 ASPGPGQPE GPRQGRVAF PLPWLPLAS HPSPLSLPH RVHAGRRDP  
301 GGPVSVPRAA AQLSPGKGA SPSPSLARS LLCTVCKVOP PTVHGSRAQ  
351 PRPLPTRYDR PSVHPGHRP PVTSTVPSRG DM

!!AA\_SEQUENCE 1.0  
P1:S02383 - probable membrane antigen Cl3 - human herpesvirus 4 (fragment)  
C:Species: human herpesvirus 4, Epstein-Barr virus  
C>Date: 01-Dec-1989 #sequence\_revision 01-Dec-1989 #text\_change 07-Feb-1997  
R:Wallis, D.; Gannon, F.  
EMBO J. 7: 1191-1196, 1988  
A:Title: The expression of novel antigens from the Epstein-Barr virus large  
internal repeat  
A:Reference number: S02381; M01D:88296424  
A:Accession: S02383  
A:Molecule type: DNA  
A:Residues: 1-57 <MOL>  
A:Cross-references: EMBL:X07816  
C:Keywords: membrane protein; surface antigen

S02383 Length: 57 February 11, 2000 15:51 Type: P Check: 5478 ..

1 RGNFCPSLCP SEEPSTGCTP EPLGPASRRP PGLRSPSPV KPKCLNGAT  
51 LGAQAE  
!!AA\_SEQUENCE 1.0  
P1:S03243 - hypothetical protein B-115 - Sulfolobus particle SSV1  
C:Species: Sulfolobus particle SSV1  
C>Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 18-Jun-1993  
R:Palin, P.; Grampp, B.; Zillig, W.; Yeats, S.; McWilliams, P.; Reiter, W.D.  
submitted to the EMBL Data Library, March 1988  
A:Reference number: S03211  
A:Accession: S03243  
A:Molecule type: DNA  
A:Residues: 1-115 <PAL>  
A:Cross-references: EMBL:X07234

S03243 Length: 115 February 11, 2000 15:51 Type: P Check: 3042 ..

1 MTEYNANSIR AKILRRKILQ LAENVYLSA SLISHTLLS YATVLRHRI  
51 LNDCEYIELY KQGRITYANI RDNARQIOL NSELEGEKFNV SGPILTKDE  
101 TPKEFGKDS LTRG  
!!AA\_SEQUENCE 1.0  
P1:S21504 - hypothetical protein (gene 29 5' region) - phage SP01  
C:Species: phage SP01  
C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Sep-1997  
R:Wilhelm, K.; Koch, T.; Rueger, W.

submitted to the EMBL Data Library, June 1991  
A:Description: deoxyuridylylate hydroxymethylase gene of phage SP01 Bacillus  
subtilis.

A:Reference number: S21504  
A:Accession: S21504  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-148 <WIL>  
A:Cross-references: EMBL:X60728; NID:914869; PID:914870

S21504 Length: 148 February 11, 2000 15:51 Type: P Check: 7782 ..

1 MNOQVVLKIG GVVVDPOPK VESKYSLE VTRKKVHNM LDIITPVGLA  
51 TKLTMALSPG ALAAGVDSA DKIRGFHDI IDVFTALAE ILMEYALTAC  
101 VLIATKKNKA GWERLKNVGY AVAGIALIPT FSEFLRWSS IVSSITF

!!AA\_SEQUENCE 1.0  
P1:S35282 - eaa protein - phage P22  
C:Species: phage P22  
C>Date: 31-Dec-1993 #sequence\_revision 02-Jun-1994 #text\_change 07-May-1999  
R:Wulff, D.L.; Ho, Y.S.; Powers, S.; Rosenberg, M.  
Mol. Microbiol. 9, 261-271, 1993  
A:Title: The int genes of bacteriophages P22 and lambda are regulated by  
different mechanisms  
A:Reference number: S35280; M01D:94018622  
A:Accession: S35282  
A:Molecule type: DNA  
A:Residues: 1-317 <WUL>  
A:Cross-references: EMBL:I06296; NID:9215287; PID:9215290  
C:Genetics:  
A:Gene: eaa

S35282 Length: 317 February 11, 2000 15:51 Type: P Check: 7758 ..

1 MTTITERIE LTVKSPLENG LTRGEOMEIA RIALASLDAE TVRYLNKFSG  
51 TCVTLEQOPN AADVAVYIP LYAAPVPER ERIRREHAW SDRTFDVG  
101 VGPLKRLSKE ALAAAPSD PLEWADMQL LMDAQRNGI SDEFTIRAMI  
151 EKLEINKTRO WPEPKDCEPR LHIKEPESV VPESCPAELP YAQKAVADL  
201 VALCWQSEV VTYTPDEKA TIMINNSGT CYQEVYALR LQALAGNP  
251 VPGGMISCS ERMNDDESK PLAITGKCL GQGNFVATYD DQGFDPYWG  
301 MEIIVSHM QLPDPPL

!!AA\_SEQUENCE 1.0  
P1:S17834 - acetyl-CoA carboxylase - phage T7  
C:Species: phage T7  
C>Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 24-Jul-1997  
R:Lopez-Castillas, F.; Kim, K.H.  
Eur. J. Biochem. 201, 119-127, 1991  
A:Title: The 5' untranslated regions of acetyl-coenzyme A carboxylase mRNA  
provide specific translational control in vitro.  
A:Reference number: S17834; M01D:92007865  
A:Accession: S17834  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-36 <LOP>

S17834 Length: 36 February 11, 2000 15:51 Type: P Check: 1383 ..

1 MLDQKGTGV QIRPLCLSR ATVGGKLTG LMSDQ  
!!AA\_SEQUENCE 1.0  
P1:A56663 - capsid protein VP1 - foot-and-mouth disease virus Asia (strain Asia  
1.63/72)



C:Species: Apthovirus Asia (foot-and-mouth disease virus Asia)  
 C>Date: 08-Sep-1995 #sequence\_revision 08-Sep-1995 #text\_change 29-Jan-1999  
 C:Accession: A56663  
 R:Sutyanarayana, V.V.; Rao, B.U.; Padayatty, J.D.  
 Indian J. Biochem. Biophys. 29, 20-24, 1992  
 A:Title: Nucleotide sequence of the cDNA and the derived amino acid sequence for the major antigenic protein of foot and mouth disease virus, type Asia 1 63/72.  
 A:Reference number: A56663; MUID:92275633  
 A:Accession: A56663  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-299 <SUR>  
 A:Cross-references: GB:S37121; NID:q249692; PID:q249693  
 A:Note: sequence inconsistent with nucleotide translation  
 A:Note: sequence extracted from NCBI backbone (NCBIN:104727, NCBI:P:104728)  
 A56663 Length: 299 February 11, 2000 15:51 Type: P Check: 6051 ..

1 MTNCCRGGL DPGFRHVLV GRVAVFIAGD ADILRLHDA GHCHRSQGT  
 51 RYALDLVNRQ FAGRQORAR RSRRCETPA GSARACRARR RSSPRHSGP  
 101 CHRASAPMSI AVCPVFSMKY ATRVSPSCAA TRPSIAKGP NAGODIAIL  
 151 RVADRRLIDE DLEELVIDD ALPIGPDDR DLAGORIGAA HAIDLARVGR  
 201 SHGSOQEGVA RGAVGQVFR KEIALAGRAA AHPHDEAGR DVRHISGPP  
 251 AGAGAYCRR GPALRRPGR ANLPFRPRAC AFSSTAASA RGMHHPPL

11AA\_SEQUENCE 1.0  
 P1:H72245 - 2,3,4,5-tetrahydropyridine-2-carboxylate  
 N:succinyltransferase-related protein - Thermotoga maritima (strain MSB8)  
 C:Species: Thermotoga maritima  
 C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 11-Jun-1999  
 C:Accession: H72245  
 R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, E.R.; Petersen, J.D.; Nelson, W.C.; Ketchum, K.A.; McDonald, L.; Usterback, T.R.; Malek, J.A.; Linher, K.D.; Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; Heidelberg, J.; Sutton, G.G.; Fleischmann, R.D.; White, O.; Salzberg, S.L.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Nature 399, 323-329, 1999  
 A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima.  
 A:Reference number: A72200; MUID:99287316  
 A:Accession: H72245  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-236 <ARN>  
 A:Cross-references: GB:AE001799; GB:AE000512; NID:q4982067; PID:q4982085;  
 A:Experimental source: strain MSB8  
 C:Genetics: TM1519  
 A:Gene: TM1519  
 H72245 Length: 236 February 11, 2000 15:51 Type: P Check: 6860 ..

1 MSELDARETI EMIAKAKKT PIYAIKGL AGIDSSSKF FODERGILF  
 51 GEYEDFRKLL EHRREKIEDY HLEVAKRNSA LPLADLTYYK ARIEPAIIR  
 101 DWEIGEGAV IMGAVINVG AVIGEGMID MNAVVGKRAI IGKKCHIGAG  
 151 AVIAGVIEPP SAKPVIEDE VLVGANAAIL EGVTVGKAV VAAGAVYTKD  
 201 VPPYTVAGV PARVIKQIDE KTKRKTIYD EIRNILE

11AA\_SEQUENCE 1.0  
 P1:E72316 - hypothetical protein - Thermotoga maritima (strain MSB8)  
 C:Species: Thermotoga maritima  
 C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 11-Jun-1999

C:Accession: E72316  
 R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, E.R.; Petersen, J.D.; Nelson, W.C.; Ketchum, K.A.; McDonald, L.; Usterback, T.R.; Malek, J.A.; Linher, K.D.; Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; Heidelberg, J.; Sutton, G.G.; Fleischmann, R.D.; White, O.; Salzberg, S.L.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Nature 399, 323-329, 1999  
 A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima.  
 A:Reference number: A72200; MUID:99287316  
 A:Accession: E72316  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-410 <ARN>  
 A:Cross-references: GB:AE001757; GB:AE000512; NID:q4981464; PID:q4981479;  
 A:Experimental source: strain MSB8  
 C:Genetics: TM0941  
 A:Gene: TM0941  
 E72316 Length: 410 February 11, 2000 15:51 Type: P Check: 6581 ..

1 MYRISFLIL ILSSLIFSK EVLENTYNT GIVSLNRTF PFETSRDLB  
 51 FKASLEFIYT PKYVGITID LPDSTLVLS NGVTIGVDVG DIRPEDVFA  
 101 ELLESTDLAF YDAPVEVE SDRVYVKS ITRKLEYLL SFEFKIEVP  
 151 RRIEGTIDVA PKPEVRSYK VMHPLIGTVY FLNVYDSDTF VILMNGREGK  
 201 LWTQTEATK LEVEVDSIG OSTSLSEAL PVSFAERSYQ SSVEFGOSLI  
 251 LPLVNVETLP STGEKITVS PEPPGVYHLI GCDSEGNFLR MEIDVRDTP  
 301 PVLSPKIE DASPRVEYI CDGREAKIP DGHVVFVKA TDFGNTSTRA  
 351 FFWVNRHRY IYERPVVY LQPRKIQIG GLSLGSLIY GWTRREEVEM  
 401 VGEVYKLEK

11AA\_SEQUENCE 1.0  
 P1:B40609 - Probable lipoprotein MlpA - Myxococcus xanthus  
 C:Species: Myxococcus xanthus  
 C>Date: 16-Feb-1994 #sequence\_revision 18-Nov-1994 #text\_change 20-Mar-1998  
 C:Accession: B40609  
 R:Martinez-Camamero, M.; Munoz-Dorado, J.; Fares-Vidal, E.; Inouye, M.; Inouye, S.  
 J. Bacteriol. 175, 4756-4763, 1993  
 A:Title: Oar, a 115-kilodalton membrane protein required for development of Myxococcus xanthus.  
 A:Reference number: A40609; MUID:93328680  
 A:Accession: B40609  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 1-236 <MAR>  
 A:Cross-references: GB:S64103; NID:q402803; PID:q402805  
 A:Note: sequence extracted from NCBI backbone (NCBIN:135644, NCBI:P:135646)  
 B40609 Length: 236 February 11, 2000 15:51 Type: P Check: 8809 ..

1 MKNIVMTAL VLVGASLLT GCNFEOPETN CFEQESPWA VKYDVVSPK  
 51 DANDDECTTT APLVELMGVY KYVNPETGAA QALRPATIA SRAIADTTT  
 101 SADQTSIGSL DTEPKDHGC HANDEAPAFV NVAASDTAA NTIRYEFTNV  
 151 RYSAVAAPG TQFTGELKYT SNGTSSVYM RAVWPAPCD TASTEPAPENC  
 201 GVSGSLNPEF AVYCPITSAT GTTGCVAPG DIPSEK

11AA\_SEQUENCE 1.0  
 P1:JC6019 - response-regulator protein Firz - Myxococcus xanthus

C:Species: Myxococcus xanthus  
 C:Date: 15-Aug-1996 #sequence\_revision 18-Oct-1996 #text\_change 07-May-1999  
 C:Accession: J66019  
 R:Rundau, K.G.; Ward, M.J.; Zusman, D.R.  
 M:J. Microbiol. 20, 645-653, 1996  
 A:Title: Identification and characterization of FrzZ, a novel response regulator necessary for swarming and fruiting-body formation in Myxococcus xanthus.  
 A:Reference number: J66019; MUID:96347136  
 A:Accession: J66019  
 A:Molecule type: DNA  
 A:Residues: 1-290 <TRU>  
 C:Cross-references: GB:U47814; NID:g1236916; PID:g1236917  
 C:Comment: This protein belongs to the response-regulator superfamily, it is involved in the chemotactic signal-transduction pathway that functions in a manner analogous in enteric bacteria, it regulates the individual cell reversal, responds to attractants and repellents in the spatial chemotaxis assay, and it is necessary for swarming and fruiting-body formation.  
 C:Genetics:  
 A:Gene: frzZ  
 C:Superfamily: response regulator homology  
 C:Keywords: phosphoprotein; signal transduction  
 F:4-114/Domain: response regulator homology <RRH>  
 F:124-162/Region: alanine-proline-rich  
 F:170-283/Domain: response regulator homology <RRH>  
 F:52/Binding site: phosphate (Asp) (covalent) #status predicted  
 F:220/Binding site: phosphate (Asp) (covalent) #status predicted  
 J66019 Length: 290 February 11, 2000 15:51 Type: P Check: 2747 ..

1 MSRVLYIDS PMLVLYTRA LTAAGYASG AODLASLEIK LAEPFSLIT  
 51 MOVNPEMG DDVVEYLKQ KYTAKLVLY SPISSEALDG KTKASGADGY  
 101 ILSGGLNAV LGGWGLIGP PALSIPTAV APAAPAAP TPPAPAPAPA  
 151 PAATGLAKA PTTGGRKPR LIYDSEMTA RITEADLYTK GFEVHVADTA  
 201 DKATILKQ QTRPDVLLD VRMPVNGEO FCRFKNSL FKGIKVLCS  
 251 GENVELORI CREAADGYI PKDAVGNLV AKELMPTGNE

!!AA\_SEQUENCE 1.0  
 P1:C70198 - conserved hypothetical protein B80788 - Lyme disease spirochete  
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
 C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 05-Jun-1998  
 C:Accession: C70198  
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, B.; White, O.; Ketchum, K.A.; Dodson, R.; Hickey, E.K.; Gwin, M.; Dougherty, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman, J.; Uitterback, T.; Watthey, L.; McDonald, L.; Artlach, P.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
 Nature 390, 580-586, 1997  
 A:Authors: Smith, H.O.; Venter, J.C.  
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.  
 A:Reference number: A70100; MUID:98065943  
 A:Accession: C70198  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-440 <TRU>  
 A:Cross-references: GB:AE001177; GB:AE000783; NID:g2688711; PID:g2688713;  
 TIGR:B80788  
 A:Experimental source: strain B31  
 C70198 Length: 440 February 11, 2000 15:51 Type: P Check: 1619 ..

1 MHFLDENIQI KIDKFKKNS LDKRNVYAF SGGADSTALL INLKYYLSNN  
 51 VIAFPAHFI RSDNEDONEI EHVGFCDLY NIALQIKKD IDIKESARL  
 101 GVSIEELAK FRYIALENAL KENGANYIAL AHNEDQIEI IMRFQGSF

1 LDGSLGIPSV NRIIRPLE VSRLEIENFL SLNNIGFEVD STNAONLYLR  
 201 NRVANNLLPA IKRVFGEYK CKIRISEBK EPADYEGKDE FPPVKGKRY  
 251 XSFDLKFLD FPKYLVFRI FKLNSGIA AKVSYALNE AKRVEINRK  
 301 NNVLKTNAA FLEKRNKTN LIFRDEKEY KPDFLLEVG KWHSLSLKI  
 351 LKYLECNA SVSRLEKCSY EFRYKFKDR LNAKRFESKF IRCNPAYML  
 401 LALNRLIGI IDLNTINLW SEKSILKIN ISLIGLKE

!!AA\_SEQUENCE 1.0  
 P1:E70167 - hypothetical protein B80542 - Lyme disease spirochete  
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
 C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 05-Jun-1998  
 C:Accession: E70167  
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, B.; White, O.; Ketchum, K.A.; Dodson, R.; Hickey, E.K.; Gwin, M.; Dougherty, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman, J.; Uitterback, T.; Watthey, L.; McDonald, L.; Artlach, P.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
 Nature 390, 580-586, 1997  
 A:Authors: Smith, H.O.; Venter, J.C.  
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.  
 A:Reference number: A70100; MUID:98065943  
 A:Accession: E70167  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-192 <TRU>  
 A:Cross-references: GB:AE001156; GB:AE000783; NID:g2688459; PID:g2688469;  
 TIGR:B80542  
 A:Experimental source: strain B31  
 E70167 Length: 192 February 11, 2000 15:51 Type: P Check: 2658 ..

1 MDNLNSINSI KOLSKFPMYK LFLIVFVLS CSSIFKEVON ISGEYKLLAK  
 51 LNEELGNET SVLIYKSIK FNINAGDSS YNFIACINL KRYVEALKL  
 101 NSIIEEDEN ILINKGYL LFRKNDLNA LIYIKLIEF APANKREALFN  
 151 IFYIHLKSD KNAKKYISK YKELNHPIS GAEEIVSIL KS

!!AA\_SEQUENCE 1.0  
 P1:B70174 - hypothetical protein B80595 - Lyme disease spirochete  
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
 C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 05-Jun-1998  
 C:Accession: B70174  
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, B.; White, O.; Ketchum, K.A.; Dodson, R.; Hickey, E.K.; Gwin, M.; Dougherty, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman, J.; Uitterback, T.; Watthey, L.; McDonald, L.; Artlach, P.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
 Nature 390, 580-586, 1997  
 A:Authors: Smith, H.O.; Venter, J.C.  
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.  
 A:Reference number: A70100; MUID:98065943  
 A:Accession: B70174  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-208 <TRU>  
 A:Cross-references: GB:AE001161; GB:AE000783; NID:g2688515; PID:g2688526;  
 TIGR:B80595  
 A:Experimental source: strain B31  
 B70174 Length: 208 February 11, 2000 15:51 Type: P Check: 8103 ..

Page 127

P1.H64586 - cgl pathogenicity island protein cag15 - Helicobacter pylori  
 (strain 26595)  
 C/Species: Helicobacter pylori  
 C/Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 16-Oct-1998  
 C/Accession: H64586  
 R/Om, J.F.; White, O.; Karlavage, A.R.; Clayton, R.A.; Sutton, G.G.;  
 Fleischmann, R.D.; Ketchum, K.A.; Klenk, H.P.; Gill, S.; Dougherty, B.A.;  
 Nelson, K.; Quackenbush, J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus,  
 B.; Richardson, D.; Dodson, R.; Khakh, H.G.; Gilex, A.; McKenney, K.;  
 Fitzgerald, L.M.; Lee, N.; Adams, M.D.; Hickey, E.K.; Berg, D.E.; Gocayne,  
 J.D.; Uitterlind, L.R.; Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman,  
 J.M.; Fujii, C.; Bowman, C.; Wathey, L.  
 Nature 388, 539-547, 1997  
 A/Nature 388, 539-547, 1997  
 A/Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.;  
 Fraser, C.M.; Venter, J.C.  
 A>Title: The complete genome sequence of the gastric pathogen Helicobacter  
 pylori.  
 A/Reference number: A64520; MUID:97334457  
 A/Accession: H64586  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-114 <TOM>  
 A/Cross-references: GB:AE000568; GB:AE000511; NID:g2313646; PID:g2313651;  
 TIGR:HP0536

H64586 Length: 114 February 11, 2000 15:52 Type: P Check: 7347 ..

1 MKRPISKQ NFLOFRSFN KKLDDYSLYI RLENISSIVY GFLVALSYSG  
 51 AGVILVYPII FLFALIKPS FRYITTYILL LVSLSIISNY YLGSANFTM  
 101 KLIIIMTQWQ NWFL

1. Nucleotide sequence of the *hlyA* gene of *Helicobacter pylori* strain 26695)  
 Pl.F64587 - gag pathogenicity island protein cag21 - *Helicobacter pylori*  
 C.Species: *Helicobacter pylori*  
 C.Date: 09-Aug-1997 #sequence, revision 09-Aug-1997 #text-change 16-Oct-1998  
 C.Accession: F64587  
 R.Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.;  
 Eichmann, R.D.; Ketchum, K.A.; Klenk, H.P.; Gill, S.; Dougherty, B.A.;  
 Nelson, K.; Quackenbush, J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus,  
 B.; Richardson, D.; Dodson, R.; Khakhria, K.; Hunkeler, A.; McKenney, K.;  
 Fitzgerald, L.M.; Lee, N.; Adams, M.D.; Hickey, E.K.; Berg, D.E.; Goddard,  
 J.D.; Utterback, T.R.; Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman,  
 J.M.; Fujii, C.; Bowman, C.; Wathey, L.  
 Nature 388, 539-547, 1997.  
 A.Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpi, P.D.; Smith, H.O.;  
 Fraser, C.M.; Venter, J.C.  
 A.Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.  
 A.Reference number: A64520; MUID:97394467  
 A.Accession: F64587  
 A.Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A.Molecule type: DNA  
 A.Residues: 1142 <TOM>  
 A.Cross-references: GB:AE000568; GB:AE000511; NID:g2313646; PID:g2313657;  
 TIGR:HP0542  
 F64587 Length: 142 February 11, 2000 15:52 Type: P Check: 2551 ...  
 1 MKTNFYKIKL LFAMCLITGM FNNAPLNADON TDIKIDISPD MAIINSVGLVS  
 51 RDOLKIETIR ETLFOKVAVL NDYNDKNVNI KFDNISIGST QPNDLGLGNA

101 MWGIQNLMS QMAGDYGPNN PEMYGAIPY SDSSIFPIIL GY  
!!AA-SEQUENCE 1.0  
P:A64649 - Helicobacter protein HP1033 - Helicobacter pylori (strain 26695)  
C:Species: Helicobacter pylori  
C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 10-Oct-1997  
C:Accession: A64649  
R:Toomb, J.F.; White, O.; Keriavage, A.R.; Clayton, R.A.; Sutton, G.G.;



```

51 ALNQAINNAR NSLFPNTKAI RDVOMALNAV KDSNKNANFE AGNGSGGIGF
101 NELSLGKYKF LQKKRIGFR HSLFFGYOLG GVGSVPGSL IAPLPYGFNT
151 DLINMTNDR RASOEYVERR VKGLSIFYKD MGRITLADT LKRASRIIR
201 KSSGLVIGME LGASTWFASN NLTPFNQVRS RTIFOLQKFE CVRFSDEYD
251 IDRGDGNLYL GGSVELGVK VPAFKVNYVS DDYGRDLXK RVSVYLYNTY
301 YNFKKH

!!AA_SEQUENCE 1.0
P1:PH0856 - mauF protein - Paracoccus denitrificans
C:Species: Paracoccus denitrificans
C>Date: 17-Aug-1992 #sequence_revision 26-Apr-1996 #text_change 07-May-1999
C:Accession: S51048; PH0856
R:van Spanning, R.J.M.; van der Palsen, C.J.N.M.; Slotboom, D.J.; Reijnders,
W.N.M.; Stouthamer, A.H.; Dulle, J.A.
Eur. J. Biochem. 226, 201-210, 1994
A:Title: Expression of the mau genes involved in methylamine metabolism in
Paracoccus denitrificans is under control of a lysR-type transcriptional
activator.
A:Reference number: S51046; MUID:95045590
A:Accession: S51048
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-277 <VAN>
A:Cross-references: EMBL:012464; NID:9558801; PID:9558804
R:Chistostedov, A.Y.; Boyd, J.; Mathews, F.S.; Lidstrom, M.E.
Biochem. Biophys. Res. Commun. 184, 1181-1189, 1992
A:Title: The genetic organization of the mau gene cluster of the facultative
autotroph Paracoccus denitrificans.
A:Reference number: PH0856; MUID:92272706
A:Accession: PH0856
A:Molecule type: DNA
A:Residues: 129-277 <CHT>
A:Cross-references: GB:M90099; NID:9150583; PID:9150584
C:Genetics:
A:Gene: mauF

PH0856 Length: 277 February 11, 2000 15:52 Type: P Check: 7474 ..

1 MVSVEDLHGL SAGQASVPCD KLFQSPSPSA TRINVLAA LAGAAGVAL
51 ASAGPOPLW AVLGAAVAG GLSTWSPCG YSSISILRPD GGLNAVAGW
101 LPTFAMHAG YGLGALMLGG LIGIGLILAG FSGFGSTALL VGLVGLAYG
151 AHQIDELRVP YPQRAQVPH DARQRPKWV IGLYGLSLG LDLYTYVOTP
201 LLYMTLAAV FTGNIAHAA IVALFNLGRF LPVAVNALPT PDYVOAMLA
251 RHOENALAD GAILTALGAG FTVLALI

!!AA_SEQUENCE 1.0
P1:19703 - trm protein - Agrobacterium tumefaciens
C:Species: Agrobacterium tumefaciens
C>Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 19-Jul-1996
C:Accession: 19703
R:Funua, C.; Burbee, M.; Minns, S.C.
J. Bacteriol. 177, 1367-1373, 1995
A:Title: Activity of the Agrobacterium T1 plasmid conjugal transfer regulator
trm is inhibited by the product of the trm gene.
A:Reference number: 19703; MUID:95173115
A:Accession: 19703
A:Status: preliminary; translated from GB/EMBL/DDDB
A:Molecule type: DNA
A:Residues: 1-102 <RES>
A:Cross-references: EMBL:U16786; NID:9571511; PID:9571512
C:Genetics:
A:Gene: trm

139703 Length: 102 February 11, 2000 15:52 Type: P Check: 4597 ..

1 MELEDANVK KVELRPILG TRGLPPTLE TITIDAIRH RLVYKABEL
51 FOALPEYTK GOACGPOHI RYEASIEBH AQMSALNTLY SLIGFIPVY
101 VN

!!AA_SEQUENCE 1.0
P1:JC2568 - mirax protein - Rhizobium meliloti (fragment)
C:Species: Rhizobium meliloti
C>Date: 13-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 14-Jul-1995
C:Accession: JC2568
R:Leach, F.; Wacks, D.B.; Signer, E.R.
Gene 148, 87-90, 1994
A:Title: Rhizobium meliloti homologs of Escherichia coli mur genes.
A:Reference number: JC2567; MUID:95011665
A:Accession: JC2568
A:Molecule type: DNA
A:Residues: 1-118 <LEA>
A:Cross-references: GB:L25875
C:Genetics:
A:Gene: mirax

JC2568 Length: 118 February 11, 2000 15:52 Type: P Check: 7036 ..

1 LOIHVPPTG ELAVILAVI GGIGFLMFNA PPAIFMGDT GSIALGASIA
51 PLAVATKHEI VMIIGLELV IETLSVITOV FWEKRTGHRV FLIGADPPL
101 REEGLDKRPQ GDPFWMQ

!!AA_SEQUENCE 1.0
P1:S27344 - hupK protein - Rhizobium leguminosarum
C:Species: Rhizobium leguminosarum
C>Date: 25-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 17-Mar-1999
C:Accession: S27344
R:Rey, L.; Hidalgo, E.; Palacios, J.; Ruiz-Arguieso, T.
J. Mol. Biol. 228, 998-1002, 1992
A:Title: Nucleotide sequence and organization of an H(2)-uptake gene cluster
from Rhizobium leguminosarum bv. viciae containing a rubredoxin-like gene and
four additional open reading frames.
A:Reference number: S27340; MUID:93108466
A:Accession: S27344
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-370 <REV>
A:Cross-references: EMBL:X52974; NID:91167855; PID:948731

S27344 Length: 370 February 11, 2000 15:52 Type: P Check: 1548 ..

1 MFLLAGTIT GIDTVSRAL ACSVAVKANR PRGLTRMFV ROPEEAPVLA
51 GGVFSICGFR QSVARLAVL AAMDAMNDE ERLGASAGL AEFIFETLRA
101 LILQPTPLP ERFADAGRH LREALASIA IISRAKAGT SRPRLAAAE
151 RLSAATALG IPGRGDTPLP ETACAAILND VEDDHVAGR RPDPLTISD
201 AEVVARLRDE AGYASLPHLS GRIETGAYA RSASAGDEA PHLARLRAR
251 IGDVRSISLQ LIALARTGDF DCASLACSGP TPAGAGYAV ECARGRLYHO
301 IEIGGGRILA AYRIAPTEM NFHPAGPVE TLLSSPGAD EAAVRSISRL
351 AVLEDPVAF EIVREADA

!!AA_SEQUENCE 1.0
P1:S28677 - hypothetical protein 4 - Rhizobium sp. (strain IC342)
C:Species: Rhizobium sp.
C>Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Sep-1997
C:Accession: S28677

```

R:Upadhyaya, N.M.; Scott, K.F.; Tucker, W.T.; Watson, J.M.; Dart, P.J.  
 Molecule: Microb. Interact. 5: 129-143, 1992  
 A:Title: Isolation and characterization of *Rhizobium* (IC3342) genes that  
 determine leaf curl induction in pigeon pea.  
 A:Reference number: S28673  
 A:Accession: S28677  
 A:Molecule type: DNA  
 A:Residues: 1-243 <UPA>  
 A:Cross-references: EMBL:M38698; NID:g152257; PID:g152262

S28677 Length: 243 February 11, 2000 15:52 Type: P Check: 6944 ..

1 MSOLVTHGQ SRVAAVFFA AEKQDLSLF DAQPERFTT AAQWEGDQ  
 51 ARHIEVEVG TLRAVRLND GRRVITGFLR PGDLGVSVK EHYLYTEAI  
 101 THVELRFRSR RRESESRARA PHLRQULFSR LCDMAAOD QWYLSRRA  
 151 EEKVAQFLIM MARGOSENRR PYIELPMTRL DVADYLGMTI ETVSRITKL  
 201 AGSGVIAVIG RHAIAITKMD ALIALADGEC DGAQRSAARY AKA

!!AA\_SEQUENCE 1.0  
 P1:S26139 - signalling protein ampd - *Citrobacter freundii*  
 C:Species: *Citrobacter freundii*  
 C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 17-Oct-1997  
 C:Accession: S26139  
 R:Kopp, U.; Wiedemann, B.; Lindquist, S.; Normark, S.  
 A:Description: Wildtype and mutant ampd genes of *Enterobacter cloacae* and  
*Citrobacter freundii*.  
 A:Reference number: S26138  
 A:Accession: S26139  
 A:Molecule type: DNA  
 A:Residues: 1-187 <KOP>  
 A:Cross-references: EMBL:Z14002; NID:g40455; PID:g40456  
 C:Genetics:  
 A:Function: regulator of the production of beta-lactamase

S26139 Length: 187 February 11, 2000 15:52 Type: P Check: 4831 ..

1 MLENGWLVYD ARHVSPPHD CRPEDEKPTL LVVHNISLPP GEGGPMWIDA  
 51 LFTGTIDPDA HPFPAEIAHL ALSADCLIRR DGEVQYVPE DKRAWHAGVS  
 101 MYGGERCND FSIQIELEGT DTPPYTDAQY EKLAVTQTL IGRYPAIADN  
 151 ITGHSIDIAP RKTDPGPAPD WSRFHMLTT SSDKEIT

!!AA\_SEQUENCE 1.0  
 P1:A48901 - signalling protein ampd - *Enterobacter cloacae*  
 C:Species: *Enterobacter cloacae*  
 C:Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 15-May-1998  
 C:Accession: A48901; S26138  
 R:Kopp, U.; Wiedemann, B.; Lindquist, S.; Normark, S.  
 A:Title: Sequences of wild-type and mutant ampd genes of *Citrobacter freundii*  
 and *Enterobacter cloacae*.  
 A:Reference number: A48901; MUID:93199292  
 A:Accession: A48901  
 A:Molecule type: DNA  
 A:Status: preliminary  
 A:Residues: 1-187 <KOP>  
 A:Cross-references: EMBL:Z14003; NID:g40905; PID:g40906  
 A:Note: sequence extracted from NCBI Backbone (NCBI:127354, NCBI:P:127356)  
 C:Genetics:  
 A:Gene: ampd

A48901 Length: 187 February 11, 2000 15:52 Type: P Check: 4831 ..

1 MLENGWLVYD ARHVSPPHD CRPEDEKPTL LVVHNISLPP GEGGPMWIDA

51 LFTGTIDPDA HPFPAEIAHL ALSADCLIRR DGEVQYVPE DKRAWHAGVS  
 101 MYGGERCND FSIQIELEGT DTPPYTDAQY EKLAVTQTL IGRYPAIADN  
 151 ITGHSIDIAP RKTDPGPAPD WSRFHMLTT SSDKEIT

!!AA\_SEQUENCE 1.0  
 P1:S40867 - ferredoxin-NADP+ reductase (Ec 1.18.1.2) - *Escherichia coli*  
 N:Alternate names: methyl viologen-resistance protein  
 C:Species: *Escherichia coli*  
 C:Date: 19-May-1994 #sequence\_revision 01-Sep-1995 #text\_change 07-May-1999  
 C:Accession: S40867; A47077; A26225; B45248; G65198; Q00073  
 R:Plunkett III, G.; Burland, V.; Daniels, D.L.; Blattner, F.R.  
 A:Title: Analysis of the *Escherichia coli* genome. III. DNA sequence of the  
 region from 87.2 to 89.2 minutes.  
 A:Reference number: S40802; MUID:93347969  
 A:Accession: S40867  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-748 <PRU>  
 A:Cross-references: EMBL:L19201; NID:g304961; PID:g305027  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October  
 1993

R:Blanch, V.; Reichard, P.; Eliasson, R.; Pontis, E.; Krook, M.; Jorvall, H.;  
 Hagbard-Ljungquist, E.  
 J.Bacteriol. 175, 1590-1595, 1993  
 A:Title: *Escherichia coli* ferredoxin NADP+ reductase: activation of E. coli  
 anaerobic ribonucleotide reduction, cloning of the gene (fpr), and  
 overexpression of the protein.  
 A:Reference number: A47077; MUID:93194782  
 A:Accession: A47077  
 A:Molecule type: DNA  
 A:Residues: 1-43, 'A', '45-150', 'T', '152-248' <BIA>  
 A:Cross-references: GB:L04757  
 A:Note: sequence inconsistent with the nucleotide translation  
 A:Note: sequence extracted from NCBI Backbone (NCBI:127345, NCBI:P:127346)  
 R:Moriyo, M.  
 J.Bacteriol. 170, 2136-2142, 1988  
 A:Title: Isolation and characterization of methyl viologen-sensitive mutants of  
*Escherichia coli* K-12.  
 A:Reference number: A26225; MUID:88198002  
 A:Accession: A26225  
 A:Status: significant sequence differences  
 A:Molecule type: DNA  
 A:Cross-references: GB:M19644  
 A:Note: original sequence reported in A26225 was incorrect due to frameshifts;  
 carboxyl-terminal segment of glpX protein was misidentified as part of mvaA  
 gene product

R:Turner, V.; Boos, W.; Sweet, G.  
 J. Bacteriol. 174, 6981-6991, 1992  
 A:Title: Molecular analysis of the glpFKX regions of *Escherichia coli* and  
*Shigella flexneri*.  
 A:Reference number: A45248; MUID:93015762  
 A:Accession: B45248  
 A:Molecule type: DNA  
 A:Residues: 1-93, 'RDSLSCKKCRNARKYGCWQVQRLALIRECN', '127-130' <TRU>  
 A:Note: sequence extracted from NCBI Backbone (NCBI:117058)  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;  
 Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,  
 J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of *Escherichia coli* K-12.  
 A:Reference number: A64720; MUID:97426617  
 A:Accession: G65198  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-248 <BLAT>  
 A:Cross-references: GB:AE000467; GB:U00096; NID:g1790356; PID:g1790359;  
 UMG:P\_b3924  
 A:Experimental source: strain K-12, substrain M6155  
 C:Genetics:

Mon Feb 14 08:07:32 2000

pir.cat

Page 131

A:Gene: fpr: mvrA  
A:Map position: 88 min  
C:Keywords: NADP, oxidoreductase

540867 Length: 248 February 11, 2000 15:52 Type: P Check: 3905 ..

1 MADWMTGKVT KYQWNTDALF SLVYHAFVLP FTAGFTKLG LEIDGENVOR

51 AYSYVNSPDN PDLEFFLYTV PDGKLSFRLA ALKPGDEVOV VSEAGFFVL

101 DEVPHCEITM MWTGTAIGP YLSILOGRD LDREKNVIV HAARYADLS

151 YELPMOLEK RYEGKRIQT VSRHETAGS LTGRIPALIE SGELESTIGL

201 PANKRSTSHVA LCGNPQWVD TOOLLKETRO MKHLRRRPQ HMTAEHYW

!!A\_SEQUENCE 1.0

P1:A64750 - hypothetical protein b0249 - Escherichia coli

C:Species: Escherichia coli

C>Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 29-May-1998

C:Accession: A64750

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;

Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,

J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: A64750

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 179 <BLAT>

A:Cross-references: GB:AE000133; GB:U00096; NID:92367099; PID:q1786443;

UMGP:b0249

A64750 Length: 79 February 11, 2000 15:52 Type: P Check: 8141 ..

1 MTOSVLLPQG PPTRRQAAV TTYYSNTLE DDOGSHERLV VDTBGRMVM

51 RANWFEPDAG EGINRYIRTS GIRDTATR

!!A\_SEQUENCE 1.0

P1:C65059 - hypothetical protein b2775 - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C>Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 14-Nov-1997

C:Accession: C65059

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;

Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,

J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: C65059

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1425 <BLAT>

A:Cross-references: GB:AE000361; GB:U00096; NID:92367160; PID:q2367161;

UMGP:b2775

A65059 Length: 425 February 11, 2000 15:52 Type: P Check: 1867 ..

1 MOHNSYRMT TLAIFSGG VSFDAVLYR IYQIPAKRM GFSNTEIGLI

51 MSTFGIAIIV LVAPOGVAD KESHRKMTS AMITIGLIGL IMATTPILMV

101 MCQIQAFAI TTILMLSVS IKAASLGDH SEQKMGWV EGRGVGWS

151 LAVETMWVS RFAPDDSTSL KTVIIYSV YIIIGLICWF EYSDNNLR

201 ANNEKOSFO LSDILAVLR STWYCSMV EGVFTYVAIL SYSINYITEM

251 YGMSLVASY MGIVINKFR ALGGLGII TYSKVSPT RVIIQLSVLG

301 LITLITALLVT NSNPQVANG IGLILLIGFT CVASRGLYNA CGEARPTSPY  
351 INGTIVGICS VIGFLPDVEV YPIIGHMODT LPAAEAVRNK WLMGMAIGM  
401 VIVFTELLFQ KIRTSADAPA MASSK

!!A\_SEQUENCE 1.0

P1:B65068 - hypothetical protein b2849 - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C>Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 14-Nov-1997

C:Accession: B65068

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;

Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,

J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: B65068

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1145 <BLAT>

A:Cross-references: GB:AE000368; GB:U00096; NID:92367165; PID:q2367167;

UMGP:b2849

A65068 Length: 145 February 11, 2000 15:52 Type: P Check: 6883 ..

1 MRYTDIEFS QIHEMYMHD IVNSDSKKRP RPLKRFLLA ENVLNQTSM

51 TLNSRYVWNV SYNVVYKSK VKNYSYRSV NDEFSLTDE INSEKRETLV

101 SSIDSLSKLV LNNLSLVLT STVRRNNRA KMWFEPSWI CIRCC

!!A\_SEQUENCE 1.0

P1:F64963 - nicotinate-nucleotide--dimethylbenzimidazole

phosphoribosyltransferase (EC 2.4.2.21) - Escherichia coli

C:Species: Escherichia coli

C>Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 12-Feb-1999

C:Accession: F64963

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;

Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,

J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: F64963

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1359 <BLAT>

A:Cross-references: GB:AE000291; GB:U00096; NID:q1788298; PID:q1788300;

UMGP:b1991

A64963 Length: 359 February 11, 2000 15:52 Type: P Check: 7029 ..

1 MQLIADLNT IPADSTAMS RAQRHIDGLL KPVSGLGLE VLAIOLAGMP

51 GLNGIRHVAK KAVLYMCADH GVABEVALS PKEVTAIOAE NMRGTTGVC

101 VLAQAGANV HVIDVGIDTA EPIPLGNR VARSGNALS APMSRQAE

151 KLIDVICT QELAKNGYTL FGVEGLGMAN TTPAAIYST ITGRDEEVY

201 GIGANLPDK LANKIDVRR AITLNPDPQ DGVDVIAKVG GFDVLGIAGY

dimethylbenzimidazole to nicotinate and

N(1)-(5-phospho-alpha-D-riboseyl)-5,6-dimethylbenzimidazole

C:Keywords: glycocylyltransferase; pentosyltransferase

```

251 MGAAACGCP VLLDGFLLSYA AALACOMSP AKPYLIPSH LSAEKARIA
301 LSHLGLPEYL NMEKRLGEGS GAALAMPITE AACAIYNNMG ELAASNYLP
351 GNTTSDLNS

!!AA_SEQUENCE 1.0
P1:T00211 - type II secretion pathway related protein etpc - Escherichia coli
C:Species: Escherichia coli
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 07-May-1999
C:Accession: T00211
R:Blattner, F.R.; Ishii, K.; Yasunaga, T.; Hattori, M.; Yokoyama, K.; Yatsudo,
H.C.; Kubota, Y.; Yamachi, Y.; Iida, T.; Yamamoto, K.; Honda, T.; Han, C.;
Ohtsuda, A.; Kasamatsu, M.; Hayashi, T.; Kuhara, S.; Shingawa, H.
DNA Res. 5, 1-9, 1998
A>Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an
enterohemorrhagic Escherichia coli O157:H7 derived from Sakai outbreak.
A:Reference number: Z14127; MUID:98290540
A:Accession: T00211
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-291 <MAK>
A:Cross-references: EMBL:AB011549; NID:d1204561; PID:d1032719
A:Experimental source: strain EHEC O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: etpc
A:Genome: plasmid pO157

T00211 Length: 291 February 11, 2000 15:52 Type: P Check: 5937

1 MLEFLSRGD KGLFKDIYL KALTPNRLC VILLIAGYQL VSVIHFMLT
51 QAAVPELSR VSAPETAVTG DQEEERFVET LFGRAPISS EGRAOETMPS
101 LSDDLSEGD LVRGILYSS VAESVAIFA HNNROFSLV GEKPSYDAT
151 ISAFSDHIY INVQKTVSL PLRYDNTERR NAYDNNMLTV GDVITODNFR
201 VESVFDMSE SAVTVNNTLS GYRLIPGKS SLFYAGLHD NDLAVSNGS
251 ELRDTROAQO INKQPELKE IKITVERDQ LYDAIVAGE N

!!AA_SEQUENCE 1.0
P1:H64726 - yabp protein - Escherichia coli
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 15-May-1998
C:Accession: H64726
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;
Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,
J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A>Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: H64726
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-216 <BLAT>
A:Cross-references: GB:AE000116; GB:U00096; NID:g1786240; PID:g1786242;
UMGP:b0056
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: yabp

H64726 Length: 216 February 11, 2000 15:52 Type: P Check: 98

1 MKVSPGAPV TLLNMRKNDI YKWSGDKMD VKMNIQRLM ETLRLHMSD
51 KQTEAYKLLF NFVNNOTGNI NASEFTGAI NENEREKFIN SLELFNKKT
101 CAKNDELIVA KGNMRVAVQT FGDIELSVTF FIKKNICIQ TLOLHKOGN

151 LGVDRKAVL PGYDMRDCYL GKTKMGSD ILYERGMNA NGVLPRTL
201 PRVTLRTVL TWIYLP

!!AA_SEQUENCE 1.0
P1:A64862 - ycgZ protein - Escherichia coli
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 13-Sep-1998
C:Accession: A64862
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;
Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,
J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A>Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: A64862
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-78 <BLAT>
A:Cross-references: GB:AE000215; GB:U00096; NID:g1787405; PID:g1787411;
UMGP:d1164
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: ycgZ

A64862 Length: 78 February 11, 2000 15:52 Type: P Check: 1943

1 MHQSVTLDS AGAIRRYFAK ANLHTQETL GEIVTEILKD GRNLSRKSIC
51 AKLICRLEHA TGEEOKHYN ALIGLFE

!!AA_SEQUENCE 1.0
P1:A64882 - ycgI protein - Escherichia coli
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 24-Oct-1998
C:Accession: A64882
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;
Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,
J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A>Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: A64882
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-262 <BLAT>
A:Cross-references: GB:AE000230; GB:U00096; NID:g1787578; PID:g1787586;
UMGP:d1326
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: ycgI

A64882 Length: 262 February 11, 2000 15:52 Type: P Check: 6021

1 MIIRICITL PYLPEPENT MIVTPRARR GAPPGETEHY GRSILGAPLI
51 WFPAPAARE SGLLIAGTIG DENSSVYTL CALRTITPSL RHHVVLGVN
101 PGCGGLGARA NANGVDLNRN FPAANKKEGE TYRWNAAE ERDVLVLTGD
151 KRGSEPTQA LCOLIHRIOF AMVVSFDPL ACIEDPRHSE LGEWLAQAFE
201 LPVTSVGYE TPGSFGSMCA DLNLICITAE FPISSDEAS EKYLFAMANT
251 LRHFKDAIR PS

!!AA_SEQUENCE 1.0
P1:S20452 - hypothetical protein X - Klebsiella pneumoniae (fragment)
C:Species: Klebsiella pneumoniae
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 09-Sep-1997
C:Accession: S20452; S21837
R:Meulenbergh, J.J.M.; Sellink, E.; Rieggman, N.H.; Postma, P.W.
Mol. Gen. Genet. 232, 284-294, 1992

```



Mon Feb 14 08:07:32 2000

pir.cat

Page 133

A:Title: Nucleotide sequence and structure of the Klebsiella pneumoniae pgq operon.  
A:Reference number: S20452; MUID:92212293  
A:Accession: S20452  
A:Molecule type: DNA  
A:Residues: 1-271 <MBU>  
A:Cross-References: EMBL:X58778; NID:943903; PID:943904  
S20452 Length: 271 February 11, 2000 15:52 Type: P Check: 9681 ..

1 SVPGDDITIL WQOLATIKOL IAHSGRLRL CLSADIERC REDKVLAMVA  
51 HIEGAGGFDG EGRDLQAFYA AGVRSIGPFM NIANREGSGV NSPFGSPDI  
101 GGTLTAAGID LIKOVNALKM QIDVSHMNEK AFMDTAHNAI SPLVATHSNA  
151 HALCPQPNL TDQGLRAIRD SGGVGVNFG NAFIRADGR DSDTLITIV  
201 RHIDYLINIM GEDHVALGSD FPGITLPEL GDVAGLPRLI NTLRASGYDO  
251 LVLDKILMEN WLRVKNWQ Q

!!AA:SEQUENCE 1.0  
P1:S01838 - nifH protein - Klebsiella pneumoniae  
C:Species: Klebsiella pneumoniae  
C:Date: 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 09-Sep-1997  
C:Accession: S01838; S01703  
R:Arnold, W.; Rump, A.; Klipp, W.; Pfeifer, U.B.; Puchler, A.  
J. Mol. Biol. 203, 715-738, 1988  
A:Title: Nucleotide sequence of a 24,206-base-pair DNA fragment carrying the entire nitrogen fixation gene cluster of Klebsiella pneumoniae.  
A:Reference number: S01836; MUID:89094839  
A:Accession: S01838  
A:Molecule type: DNA  
A:Residues: 1-2220 <ARN>  
A:Cross-References: EMBL:X13303; NID:943820; PID:943826  
R:Baynon, J.; Cannon, M.; Buchanan-Wollaston, V.; Ally, A.; Setterquist, R.; Dean, D.; Cannon, F.  
Nucleic Acids Res. 16, 9860, 1988  
A:Title: The nucleotide sequence of the nifH, nifY, nifX and nifW genes of K. pneumoniae.  
A:Reference number: S01702; MUID:89041575  
A:Accession: S01703  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-169, 171-217, 'GLTGTTMKNAFNPCEFLVAR' <BEY>  
A:Cross-References: EMBL:X12595  
C:Genetics: nifY  
A:Gene: nifY

S01838 Length: 220 February 11, 2000 15:52 Type: P Check: 7149 ..

1 MSDNDLEWR MLFQSLPD LQPAQIVWL ADESGETLP ERLATLQPO  
51 LAASFPSATA VMSPARMSRV MASLOGALPA HLRIVPAOR TPQLIAFCS  
101 QDGVLINHF GQGRLEFIYA EDEQGWLDY LRRYPSPHQ OEANEVRARL  
151 IEDQLLPCQ EIGGPAAR IRHRIHPKA QPGTIIQAC EAINTLIAGR  
201 LPPWLAKRLN RDNPLEERVE

!!AA:SEQUENCE 1.0  
P1:S22619 - hypothetical protein - Salmonella choleraesuis  
C:Species: Salmonella choleraesuis  
C:Date: 22-Nov-1993 #sequence\_revision 26-May-1995 #text\_change 09-Sep-1997  
C:Accession: S22619  
R:Brown, P.K.; Komana, I.K.; Reeves, P.R.  
Mol. Microbiol. 6, 1385-1394, 1992  
A:Title: Molecular analysis of the rfp gene cluster of Salmonella serovar muenchen (strain M67): the genetic basis of the polymorphism between groups C2 and B.  
A:Reference number: S22613; MUID:92349966

A:Accession: S22619  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-336 <BRO>  
A:Cross-References: EMBL:X61917; NID:947004; PID:947011  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library. //  
S22619 Length: 336 February 11, 2000 15:52 Type: P Check: 7734 ..

1 MNKKVIMDI SWSNKGIGR FTDEISKLIC DISKEELYR CASPLAPGL  
51 AVNIFLRKKT DVFLEPGYIP PLFCSKRFII TIHDLNHLI NDNSLFRRL  
101 FYNFIRKQC RKAKITTS NFSKERYAV SGVNPKNIVY VANGVSLFN  
151 ADVKPLNIGY KYLLCVGNRK THNKECVIS AFAKADIPS IKLVGTGIPC  
201 NDLEKLIQH GLSERVKEFG EVSEKDLPSL YKSLGLVFP SLXEGFGLPV  
251 VEGMACGIPV LTLSTLSLPE VAGDAAILVD PLSEDAITKG ISRLNDBEL  
301 RKHLIQKGLL RAKRFNNQNV VSEIEKVLTE AGDGNK

!!AA:SEQUENCE 1.0  
P1:A53302 - hemF 5'-region hypothetical protein - Salmonella typhimurium  
C:Species: Salmonella typhimurium  
C:Date: 08-Sep-1995 #sequence\_revision 08-Sep-1995 #text\_change 09-Sep-1997  
C:Accession: A53302  
R:Xu, K.; Elliott, T.  
J. Bacteriol. 175, 4990-4999, 1993  
A:Title: An oxygen-dependent coproporphyrinogen oxidase encoded by the hemF gene of Salmonella typhimurium.  
A:Reference number: A53302; MUID:93352403  
A:Accession: A53302  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-289 <XNA>  
A:Cross-References: GB:U19503; NID:9310640; PID:9310641

A53302 Length: 289 February 11, 2000 15:52 Type: P Check: 5663 ..

1 MSTRKLKTL TSRROYLKTG LALITLSGMS HAAVAEETLK TSNHSKPKT  
51 KKTGSKRLVM LDPGHGIDT GAIRNGSQE KHVVALAIKN VRALIRNHGI  
101 DARLRTGDT FIPLYDRVEI AHKGDALFM SIHADGETNP KAAGASVEAL  
151 SNRGAASSMA KYLSERENRA DEVAGKATD RDHLLOQVLF DLYGTDLIKN  
201 SLTGSHTLK KIRPIKLS RTEQAALFV LKSPSIPSVL VETSFITNPE  
251 EERLIGTTAF ROKIATAIAN GIISYFWFD NOKAHTTKR

!!AA:SEQUENCE 1.0  
P1:S23906 - hypothetical protein 1 - Shigella flexneri  
C:Species: Shigella flexneri  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Sep-1997  
C:Accession: S23906  
R:Truniger, V.; Sweet, G.; Boos, W.  
Submitted to the EMBL data library, March 1992  
A:Description: glpX, a new glp gene is located at the end of the glpFK operon of Escherichia coli  
A:Reference number: S23905  
A:Accession: S23906  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-135 <TRU>  
A:Cross-References: EMBL:Z11766; NID:947027; PID:947029

S23906 Length: 135 February 11, 2000 15:52 Type: P Check: 4054 ..

1 MADWTGKVT KQONWTDALF SLTVHAPVLP FTAQGFITKG LBLDERVQR

```

51 ASIVNSPDN PDLEFYLTV PDGKISPLA ALKPDDEVY VSEOLASLEW
101 MKYRTAKRYG CWOPVORLAT IYFCNLGKD LDRFK

!!AA_SEQUENCE 1.0
P1:T14663 - histone H5 like protein - Yersinia pestis plasmid pMT1
C:Species: Yersinia pestis
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T14663
R:Hu, P.; Elliott, J.; McCreedy, P.; Skowronski, E.; Garnes, J.; Kobayashi, A.;
Carrano, A.V.; Brubaker, R.; Garcia, E.
Submitted to the EMBL Data Library, March 1998
A:Description: Structural organization of virulence determinants in three
Yersinia pestis plasmids.
A:Reference number: 218168
A:Accession: T14663
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-99 <HUP>
A:Cross-references: EMBL:AF053947; NID:g2996286; PID:g2996296; PIDN:AAC13176.1
C:Genetics:
A:Genome: plasmid pMT1

T14663 Length: 99 February 11, 2000 15:52 Type: P Check: 2204 ..

1 MRATYEGKN VSPSAVSSR NDLRADIVE TAAPIVPMK RGTAEVEAKP
51 VQRTFEELE STADEGIALG LROIGNOIGV KAKGIVEMTE GILKAOGE

!!AA_SEQUENCE 1.0
P1:T14952 - hypothetical protein - Yersinia pestis plasmid pMT-1
C:Species: Yersinia pestis
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T14952
R:Findler, L.E.; Plano, G.V.; Burland, V.; Mayhew, G.F.; Blattner, F.R.
Article: Immun. 66, 5731-5742, 1998
A:Title: Complete DNA sequence and detailed analysis of the Yersinia pestis
KRM5 plasmid encoding murine toxin and capsular antigen.
A:Reference number: Z18268; MUID:99043898
A:Accession: T14952
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-144 <LIN>
A:Cross-references: EMBL:AF074611; NID:g3883003; PID:g3883036; PIDN:AAC82696.1
C:Genetics:
A:Gene: Y1036
A:Genome: plasmid pMT-1

T14952 Length: 144 February 11, 2000 15:52 Type: P Check: 2530 ..

1 MKPAIRLLE POFLYTGIL CGIQFPGIS VAEIRFIDQ RITASMRATT
51 VEGKVVPSA AYSSRDLTA DIVERAARD IVPKRGTA EAKKPVQRT
101 REELSLADC EGIALGRQIG NOIGYAKKI VEMTEGILKA OGGE

!!AA_SEQUENCE 1.0
P1:S70883 - hypothetical protein 3 - Vibrio cholerae
C:Species: Vibrio cholerae
C>Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 17-Mar-1999
C:Accession: S70883
R:Comstock, L.E.; Johnson, J.A.; Michalski, J.M.; Morris Jr., J.G.; Kaper, J.B.
Mol. Microbiol. 19, 815-826, 1996
A:Title: Cloning and sequence of a region encoding a surface polysaccharide of
Vibrio cholerae O139 and characterization of the insertion site in the
chromosome of Vibrio cholerae O1.
A:Reference number: S70876; MUID:96417860
A:Accession: S70883
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-227 <COM>
A:Cross-references: EMBL:U47057
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, January

1996
S70883 Length: 227 February 11, 2000 15:52 Type: P Check: 2492 ..

1 MARLFIWNL SRVKKRIISV LVDAFFIFS FYSAYWVRG NVEIHSDSI
51 PYLLATIVV TILSTRLG YRAIRYITE HALAVSVGT LISASVALA
101 AFYDAPYPR SLPIIYTEL CLICGSRLI VRLVGLNG KGRKVLIVG
151 AGSAGROLAI ALRNSENVYR AGFIDNKT LENTVINGMV HDVSAATLV
201 DKYDVTOILL AVPSARALS SDCAKI

!!AA_SEQUENCE 1.0
P1:S5442 - membrane associated protein 17.9K - Vibrio cholerae
C:Species: Vibrio cholerae
C>Date: 06-Sep-1996 #sequence_revision 13-Mar-1997 #text_change 09-Sep-1997
C:Accession: S5442
R:Johnson, G.; Lebens, M.; Holmgren, J.
Mol. Microbiol. 13, 109-118, 1994
A:Title: Cloning and sequencing of Vibrio cholerae mannose-sensitive
haemagglutinin pilin gene: localization of mshA within a cluster of type 4
pilin genes.
A:Reference number: S54441; MUID:95075288
A:Accession: S5442
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-150 <COM>
A:Cross-references: EMBL:X77217; NID:g496755; PID:g496757
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, January
1994

S54442 Length: 150 February 11, 2000 15:52 Type: P Check: 6509 ..

1 MKIGERVRF VLMFVLVYL LTFMSVWRS MESDLHTAL EMTRLOIDR
51 ANTYQEWTL QGRPALLOIE QAEIPMOHGM VEPKIDQVD CERVELFLP
101 DRKVLDWLR VTDLORANGY OCRVQGRDV QPDVELKVRV FAINASFWR

!!AA_SEQUENCE 1.0
P1:G64102 - phosphatidate cytidyllyltransferase homolog - Haemophilus influenzae
(strain Rd KW20)
C:Species: Haemophilus influenzae
C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 24-Oct-1997
C:Accession: G64102
R:Flaischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.;
Kerlavage, A.R.; Buit, C.J.; Tomb, J.F.; Dougherty, B.A.; Merrick, J.M.;
McKenney, K.; Sutton, G.; Fitzhugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.F.; Phillips,
C.A.; Spriggs, T.; Hedblom, E.; Cotton, M.D.; Utterback, T.R.; Hanna, M.C.;
Nguyen, D.T.; Sauder, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.;
Futmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.;
Venter, J.C.
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.
A:Reference number: A64000; MUID:95350630
A:Accession: G64102
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-288 <TIGR>
A:Cross-references: GB:U32773; GB:I42023; NID:g1573932; PID:g1573940;
TIGR:H10919

G64102 Length: 288 February 11, 2000 15:52 Type: P Check: 8973 ..

1 MKQRYLSAI VLIAYLALC FLTFPEYAL AAGVAIIGI WENQFARLK
51 QPIRFEVTT FVGVFILML YTEGNYLDAG RYFEOHOLL LINAVSWGL

```

```

101 ALLVIVSR SAKFMSKNPL LQLFAFSTL IPFVAGVRL RLEHYTHDPY
151 HGLLELLXVF ILVMADSGA YFSGRAFGR KLAPKVSPOK SMEGIGGLI
201 TALVIAEFPI HFSNNLVGD ENIGFILLS VATVAISVIG DLTESMKRE
251 SGVKDSOLI PGHGVLDRI DSLTAAPFF SYFFFFVL

!!AA_SEQUENCE 1.0
P1:H64157 - sufl protein homolog HI0733 - Haemophilus influenzae (strain Rd
KW20)
C:Species: Haemophilus influenzae
C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 31-Oct-1997
C:Accession: H64157
R:Reischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.;
Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.; Dougherty, B.A.; Merrick, J.M.;
McKenney, K.; Sutton, G.; Fitzhugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, F.; Phillips,
C.A.; Spriggs, T.; Hedblom, E.; Cotton, M.D.; Uterback, J.R.; Hanna, M.C.;
Nguyen, D.T.; Saudek, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.;
Fuhmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.;
Venter, J.C.
A>Title: Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd
A:Reference number: A64000; MUID:95350630
A:Accession: H64157
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1311 <DIGR>
A:Cross-references: GB:U32756; GB:I42023; NID:91573729; PID:91573736;
TIDR:HI0733

H64157 Length: 311 February 11, 2000 15:52 Type: P Check: 4

1 MPRLSRQQL KTAIASTALS TVPAPLLAS REKLVVPLI EVRRGREIVL
51 TMOETNPPLD GSHNVTWGF NGNYLGPTIK IKSGSFATKN YHNNLPQSA
101 LSIQIGLQAG ELFGGARVYL KKGSMAPIV PIEOPASQW YRSATILANSA
151 YQYRGLAGM WLEDEQSLK ANLPNKYGD DIPILIDME FNNQGLDFK
201 QNOHFVGNR LLVNGIEAPY LDVARGWIRL RLINSLARA YDLALDDQE
251 MLINADLGF LPRKSVKSL VLSPEERAI LVNMKLTLY LSLAEVAVAC
301 TNRKRYVLR R

!!AA_SEQUENCE 1.0
P1:H71692 - hypothetical protein RP358 - Rickettsia prowazekii
C:Species: Rickettsia prowazekii
C>Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 07-May-1999
C:Accession: H71692
R:Anderson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichert-Ponten, T.;
Randermark, U.C.M.; Podowski, R.M.; Naeslund, A.K.; Eriksson, A.S.; Winkler, H.H.;
Kurand, C.G.
Nature 396, 133-140, 1998
A>Title: The genome sequence of Rickettsia prowazekii and the origin of
Mitochondria
A:Reference number: A71630; MUID:99039499
A:Accession: H71692
A>Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-405 <AND>
A:Cross-references: GB:AJ235271; GB:AJ235269; NID:93860788; PID:e1342662;
PID:93860918
A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: RP358

H71692 Length: 405 February 11, 2000 15:52 Type: P Check: 1607

1 MOYLILSLIF LIPSLGMLTG LSIATVAEP LLSIIITGPI SFIOKQEPNK
51 KFFIKLPFR GFDDIFCIIN KIKTELFETA WCFISCLAV HPINSLVTF
101 KVFYLLFRF VSNATVFOV LYIKNSLILG IITAILLFEI EYSSHGLFR
151 MKTHFGIYM LDRGCLLSI TTVAIILF SNGHINSEI LYIVLYLIS
201 ISDSLASFEG FSIGGIIFIL ARLIKTIFK LITISLITGS LLEPVIARQI
251 DPONISEKYL ATOPSAHLR FIVHFVANKI IIRPIGYGF ASSXYIEFGD
301 NAMIDYRGEK LHPPLPHPN NILQITLLEL ILGLAFLCL VYKIKLEIDN
351 IKVSNFRAS YSCFINTYII GMISINWOT WMLISGIML VLMLLVKPD
401 IIIDN

!!AA_SEQUENCE 1.0
P1:E72060 - C1470 hypothetical protein - Chlamydia pneumoniae (strain CWL029)
C:Species: Chlamydia pneumoniae
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Apr-1999
C:Accession: E72060
R:Kalan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.;
Grimwood, J.; Davis, R.W.; Stephens, R.S.
Nature Genet. 21, 385-389, 1999
A>Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: E72060
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-252 <ARN>
A:Cross-references: GB:AE001643; GB:AE001363; NID:94376876; PID:94376883
A:Experimental source: strain CWL029
C:Genetics:
A:Gene: CP0589

E72060 Length: 252 February 11, 2000 15:52 Type: P Check: 2665

1 MOICVIGVL RSRPLGRNHT LITLFTPEGL FTFKQGOOT LQCDYRETLV
51 PISIGKYLTH RNSGRPLKLT HGDLNLFEN IKQYVALLFA SGKMIQALLA
101 SQMKRPSHK LPSLPLNPLH RIPSNSPER FPAIVYLLK QYSGILDLP
151 ACSLCKASLP YACRYQGHK LCKKHQKQA ISIEKEEEOI LQALIHAKOF
201 SELIAIEFP IAIKRIEYL FDSIQEERKS ERNSSEDPYH EILRLSVYH
251 PY

!!AA_SEQUENCE 1.0
P1:F72031 - C1578 hypothetical protein - Chlamydia pneumoniae (strain CWL029)
C:Species: Chlamydia pneumoniae
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Apr-1999
C:Accession: F72031
R:Kalan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.;
Grimwood, J.; Davis, R.W.; Stephens, R.S.
Nature Genet. 21, 385-389, 1999
A>Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: F72031
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-493 <ARN>
A:Cross-references: GB:AE001643; GB:AE001363; NID:94377118; PID:94377121
A:Experimental source: strain CWL029
C:Genetics:
A:Gene: CP0809

F72031 Length: 493 February 11, 2000 15:52 Type: P Check: 6017

```

1 MSISSSGCD NQKINMGVL TSTPQGVPOO DKLSCENKQ IQOTROCKNT  
51 EKHSDATLNG ASGKDKTSST TKETAPQOQ VAAGRESSES OKGADTGV5  
101 GAAATTAANT ATKIAMOTSI EEAASKMEST LESLOSLSAA OKMEVEAVV  
151 AALSCKSSGS AKLETPPELPK PGVTPRSEVI EIGLALAAAI QTIGEAIRKA  
201 LSNVASTQAO ADQTNKLGLE KQAIKIDKER EYEQEMKAAE OKSKDEGTM  
251 DTVNTVMIV SVAITVISIV AAFTGAGI AGLAAGAAGV AAAAGNAGA  
301 AAATVATQI TVQAVVQAVK QAVITAVROA ITAIAKAVK SGIAKIFKTL  
351 VAAIAKAIKSK GISKVFAKGT QMIAKNPPKL SKVISLSLAK WATVGVGVV  
401 AAPALGKIM QMOLEMQON VAOFOKEVGR LOAADMISM FTQWQOQASK  
451 TISKOTGESN EMTOKATKLG AQILKAYAAI SGAIAGAKRT NNF

!!AA\_SEQUENCE 1.0  
P1:S61492 - hypothetical protein 2 - Chlamydia psittaci  
C:Species: Chlamydia psittaci  
C>Date: 27-Apr-1996 #sequence\_revision 13-Mar-1997 #text\_change 09-Sep-1997  
C:Accession: S61492  
R:Rockey, D.D.; Heinzen, R.A.; Hackstadt, T.  
Mol. Microbiol. 15, 617-626, 1995  
A:Title: Cloning and characterization of a Chlamydia psittaci gene coding for a protein localized in the inclusion membrane of infected cells.  
A:Reference number: S61491; MUID:95302975  
A:Accession: S61492  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-270 <ROC>  
A:Cross-references: EMBL:L35036; NID:g516598; PID:g516600

S61492 Length: 270 February 11, 2000 15:52 Type: P Check: 5418 ..

1 MSRYSSNKKH HSRKTERNS TSWEPIAEDY HKIVGEGHY YKVEVILPKL  
51 LPLMLQSED SLVDIGCGQG ILERAIPKEC GYLGLDISPS LISTARRLRK  
101 SRDHEKIOD LTKRLVLETP QSFSAVAAIL SLQNNETPER AIKNTSKLLN  
151 DGGREFMVLN HPCFRIPRVS SMHYDEDKL LSRKIDRYLS KITVSIYANP  
201 GKQSSSSIS FHPFLSYWTQ ALSKYGFVIE NMEWISPKK SIGAHAKAEN  
251 LCREPPLFL MISCIKTHKN

!!AA\_SEQUENCE 1.0  
P1:JC5205 - sulfur-rich protein - Chlamydia psittaci  
C:Species: Chlamydia psittaci  
C>Date: 20-Feb-1997 #sequence\_revision 27-Feb-1997 #text\_change 07-May-1999  
C:Accession: JC5205  
R:Hsieh, R.; Bayoill, P.M.  
Gene 176, 135-162, 1996  
A:Title: Sequence analysis of the omp2 region of Chlamydia psittaci strain GPIC: Structural and functional implications.  
A:Reference number: JC5203; MUID:97075924  
A:Accession: JC5205  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-160 <HSI>  
A:Cross-references: GB:U41759; NID:g1783376; PID:g1783383  
A:Experimental source: strain GPIC  
C:Genetics: srp  
A:Gene: srp

JC5205 Length: 160 February 11, 2000 15:52 Type: P Check: 8587 ..

1 MTLGVENSES GVIDLIKPGI DDVAKNETVO VTLVNSVLGW CKAHIYDPK

51 TSKIQASAF QIMVUVGI LLIAGLALTE VLQQLGKNA FLFLIYAVIG  
101 LVKLITTSVF MEKPIPEKM RICKRLIAT EDIIDDQIN QSNITTTES  
151 SDVTNTATQS

!!AA\_SEQUENCE 1.0  
P1:A71529 - hypothetical protein CT324 - Chlamydia trachomatis (serotype D, strain UW3/Cx)  
C:Species: Chlamydia trachomatis  
C>Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 21-Nov-1998  
C:Accession: A71529  
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Tatusov, R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.  
Science 282, 754-759, 1998  
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis.  
A:Reference number: A71570; MUID:99000809  
A:Accession: A71529  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-303 <ARN>  
A:Cross-references: GB:AE001305; GB:AE001273; NID:g3328737; PID:g3328742  
A:Experimental source: serotype D, strain UW-3/Cx  
C:Genetics: CT324  
A:Gene: CT324

A71529 Length: 303 February 11, 2000 15:52 Type: P Check: 4189 ..

1 MYKAAHPHS SISGALPLHS SRGAYHDSL GDLSPISLT PPAISVLTPE  
51 PAQSKRVQA VANTIGDFLQ RMWKHLLEYI LWSGLILICH SSVALVLSIW  
101 LGIGGAGV LGISANFLD KENKYPHLNS LMNTINNGLO QUDPNTROV  
151 LLAIVIASIS ALIYASPOAI GFILGAFGH QTSSLAIVYG RFKSGGCIYA  
201 DQELDKQEK RIRQAIMQCR LIRNOMIILOR RLDPLQKRAH RPOQSCITTI  
251 RPLNRIISLP EMEHEILK KPOVVIAMD NKIAQLNOSL VOLYDRPTRV  
301 IEQ

!!AA\_SEQUENCE 1.0  
P1:C71569 - probable ribonuclease HII - Chlamydia trachomatis (serotype D, strain UW3/Cx)  
C:Species: Chlamydia trachomatis  
C>Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 21-Nov-1998  
C:Accession: C71569  
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Tatusov, R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.  
Science 282, 754-759, 1998  
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis.  
A:Reference number: A71570; MUID:99000809  
A:Accession: C71569  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-300 <ARN>  
A:Cross-references: GB:AE001275; GB:AE001273; NID:g3328388; PID:g3328396  
A:Experimental source: serotype D, strain UW-3/Cx  
C:Genetics: rnhb\_1  
A:Gene: rnhb\_1

C71569 Length: 300 February 11, 2000 15:52 Type: P Check: 1406 ..

1 MSSFVSQLS PSLSILREQ LEKKGFTISI PPHVTFQGRS PTVSCVYQS  
51 KTIIVQGRKT QEFYEFLEP EILQTFSSON VOQDLRSRIG VDSSGCGDF  
101 GPLCTAGIYA SSPAIEALY KTSICDSKLI PDAIILSLAQ NISLCAKCY  
151 ITLPEKINA LYANFONLNS LIAWTHATII DNLAPHPAGA VFAISDQFAS

201 SERVULONAV KKCSDIELLO RHRAEDOVV AAASILAREA FLSSIHLES  
251 OYOTRIKKA SGRVORAKE ILHNGOVVL EKVCTHFTI FNEVLGSGNQ

11AA\_SEQUENCE 1.0  
P1:B53203 - hypothetical protein 2 - Desulfovibrio vulgaris (strain Miyazaki)  
C:Species: Desulfovibrio vulgaris  
C>Date: 19-Oct-1995 #sequence\_revision 19-Oct-1995 #text\_change 09-Sep-1997  
C:Accession: B53203  
R:Kikamura, M.; Kojima, S.; Ogasawara, K.; Nakaya, T.; Sagar, T.; Niki, K.; Miura, K.; Akutsu, H.; Kumagai, I.  
J. Biol. Chem. 269, 5566-5573, 1994  
A:Title: Novel FMN-binding protein from Desulfovibrio vulgaris (Miyazaki F).  
Cloning and expression of its gene in Escherichia coli.  
A:Reference number: A53203; MIM:94164898  
A:Accession: B53203  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-43 <KIT>  
A:Cross-references: GB:D21804; NID:g425735; PID:d1005361; PID:g476038  
B53203 Length: 43 February 11, 2000 15:52 Type: P Check: 1180

1 MTVGLANPQ ANPDRHARY ARLADPAGPA SRSLPDMPG RAA

11AA\_SEQUENCE 1.0  
P1:S17812 - hypothetical protein 192 - Rhodobacter capsulatus  
C:Species: Rhodobacter capsulatus  
C>Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 09-Sep-1997  
C:Accession: S17812  
R:Burke, D.H.; Albert, M.; Armstrong, G.A.; Hearst, J.E.  
Submitted to the EMBL Data Library, November 1991  
A:Description: The complete nucleotide sequence of the 46 kb photosynthesis gene cluster of Rhodobacter capsulatus.  
A:Reference number: S17803  
A:Accession: S17812  
A:Molecule type: DNA  
A:Residues: 1-192 <EMB>  
A:Cross-references: EMBL:211165; NID:g46097; PID:g46111  
S17812 Length: 192 February 11, 2000 15:52 Type: P Check: 2127

1 MDLLEDFERA AHVPEENAT HYPEARQI GAAMSDBRG FAQVTAISR  
51 LQELHALQI LVTADSVCA NGATVLIIV PEGHTIAL IVAMELRRG  
101 VSVRIYFAPG LSDLSRLMAT TREDALITV GSDRVEICA KIVKTISSLT  
151 KGRMVAIGG AIVSORAEL ARTGADLVTN DLSVISEFS LV

11AA\_SEQUENCE 1.0  
P1:S22631 - petp protein - Rhodobacter capsulatus  
C:Species: Rhodobacter capsulatus  
C>Date: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 09-Sep-1997  
C:Accession: S22631; S21001  
R:Tokito, M.K.; Daidal, F.  
Mol. Microbiol. 6, 1645-1654, 1992  
A:Title: petp, located upstream of the fbcFBC operon encoding the cytochrome bc(1) complex, is homologous to bacterial response regulators and necessary for photosynthetic and respiratory growth of Rhodobacter capsulatus.  
A:Reference number: S22631; MIM:92356828  
A:Accession: S22631  
A:Molecule type: DNA  
A:Residues: 1-166 <TOX>  
A:Cross-references: EMBL:212113; NID:g49287; PID:e49248; PID:g1333802  
C:Genetics:  
A:Gene: petp  
C:Keywords: DNA binding; transcription regulation  
S22631 Length: 166 February 11, 2000 15:52 Type: P Check: 1875

1 MADTGAPGE TLFTIDEQL KGIEMAFPA YRGFTADPDR IIDQDIYGA

51 HHRAHFINR EPLVITLLI SVLGVTKSL NREVLITLDD GLVESRVGR  
101 DKRRHHLT EKVGLIERL SEAGRVMPA AVRAQPOAV AGFROYLEAM  
151 MDPAMRRHQ MLKDAE

11AA\_SEQUENCE 1.0  
P1:B58883 - transcription negative regulator Chr - Rhodobacter sphaeroides  
N:Alternate names: cobaltic-heme resistance protein Chr  
C:Species: Rhodobacter sphaeroides  
C>Date: 16-Oct-1998 #sequence\_revision 16-Oct-1998 #text\_change 16-Oct-1998  
C:Accession: B58883; A56261  
R:Donohue, T.J.; Newman, J.  
Submitted to Genbank, November 1996  
A:Reference number: A58883  
A:Accession: B58883  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-213 <DON>  
A:Cross-references: GB:U1283; NID:g1654106; PID:g1654107  
A:Note: sequence revision of A56261 into two ORFs. This revision submission is not cited in Genbank entry RSU11283, release 106.0  
R:Schlike, B.A.; Donohue, T.J.  
J. Bacteriol. 177, 1929-1937, 1995  
A:Title: Chr positively regulates transcription of the Rhodobacter sphaeroides cytochrome c-2 gene.  
A:Reference number: A56261; MIM:95238260  
A:Accession: A56261  
A:Molecule type: DNA  
A:Residues:  
A:Cross-references: GB:U1283; NID:g1654106  
-213 <SCH>  
A:Comment: This negative regulator of transcription inactivates the upstream encoded RpoE (see PIR:A58883).  
C:Genetics:  
A:Gene: Chr  
C:Keywords: transcription factor  
B58883 Length: 213 February 11, 2000 15:52 Type: P Check: 818

1 MTRHVSDA LITAYAGTL SEAFSLVAT HSLCDECR RAGALDAVGS  
51 SIMEETAPVA LSEGLASVM AQDROIQP APARADPRA PABLADYGR  
101 RLEDVRWRL GGVROAILP TGGEAIALRL WIPGQANPD HGRIGLETL  
151 VLOGAFRDET DRFGAGDIEI ADELEHTPV AERGIQICL AATDAPLREN  
201 SFLPKVQPF FRI

11AA\_SEQUENCE 1.0  
P1:S36981 - hypothetical protein 7 (atpC 3' region) - Synechococcus sp. (PCC 6716)  
C:Species: Synechococcus sp.  
A:Variety: PCC 6716  
C>Date: 22-Jan-1994 #sequence\_revision 12-Apr-1996 #text\_change 07-May-1999  
C:Accession: S36981; S31905  
R:van Walraven, H.S.; Lutter, R.; Walker, J.E.  
Biochem. J. 294, 239-251, 1993  
A:Title: Organization and sequences of genes for the subunits of ATP synthase in the thermophilic cyanobacterium Synechococcus 6716.  
A:Reference number: S36981; MIM:93371369  
A:Accession: S36981  
A:Molecule type: DNA  
A:Residues: 1-201 <VAN>  
A:Cross-references: EMBL:X70433; NID:g49234; PID:g49237  
A:Experimental source: PCC 6716  
S36981 Length: 201 February 11, 2000 15:52 Type: P Check: 4069

1 MYALQPRGY EIQILAPCGS RLPVPAIEE IPGALQVPAQ HORRQPIYV  
 51 PMSNVLANLW ARVROLOQGY DLIVFAVDM LPFLYFPMLS RVAVHVSMA  
 101 SISEVMDQAI ATVIDQYPS IGVYTRQQA TPEFGDRVCV LSGGSLISLY  
 151 EFCADPDVVL CWLGRIAPER GLEDAVAAN VTRTPKING QMDDVDYMR  
 201 I

!!AA\_SEQUENCE 1.0  
 P1:S74634 - heme binding protein precursor ycf10 - *Synechocystis* sp. (strain PCC 6803)  
 N:Alternate names: hypothetical protein s111685  
 C:Species: *Synechocystis* sp.  
 A:Variate: PCC 6803  
 C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 21-Aug-1998  
 C:Accession: S74634  
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hikosawa, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.; DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.  
 A:Reference number: S74322; MUID:97061201  
 A:Accession: S74634  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1593 <RAN>  
 A:Cross-references: EMBL:D90900; GB:AB001339; NID:g1651768; PID:d1017519; PID:g1651693  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 C:Genetics:  
 A:Gene: ycf10

S74634 Length: 393 February 11, 2000 15:52 Type: P Check: 2111

1 MLPKRLGSP NHRFHWLVV QALALDNAQ VSAQAISIE KTYFDGKRA  
 51 PDSRGVNTY NYKSTLERE LKIKFNLAR FNATNLNNS EDPQIEAII  
 101 IVKDAIEKI IAKYRESEI VEDSOSAE INNISPKRN FENIQRLSA  
 151 QDEKRIIQL RNRLQRIA TFLIVLFI PLTVQILTN LYPAPLVQHF  
 201 RVDIVAMEKI HYOETIEHY FEEFARYET LEIKOLSEN OPLNOEKIHQ  
 251 ELKKAEELI RQATNSQOG IVNLADLAD LVAFLVLIY FNGKSLITQQ  
 301 YLSQSFALN DITKVFIFIL LDMFVGFHS AHGMEVYLEN LSHGGLPEN  
 351 RHAVYIFIAI VPFVLSLFR LLINFTYRQ SPFSVAILEK MQQ

!!AA\_SEQUENCE 1.0  
 P1:S76728 - hypothetical protein - *Synechocystis* sp. (strain PCC 6803)  
 C:Species: *Synechocystis* sp.  
 A:Variate: PCC 6803  
 C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 21-Aug-1998  
 C:Accession: S76728  
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hikosawa, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.; DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.  
 A:Reference number: S74322; MUID:97061201  
 A:Accession: S76728  
 A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
 A:Residues: 1187 <RAN>  
 A:Cross-references: EMBL:D90916; GB:AB001339; NID:g1653715; PID:d1019373; PID:g1653729  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 C:Genetics:  
 A:Start codon: GTG

S76728 Length: 187 February 11, 2000 15:52 Type: P Check: 9759

1 MLKRSYQILS GVLAAALG FTTPAQAEPP QKGNILISQ LPGHMTLPD  
 51 PEQMSFDGLA IGRVAGKVS IIQVELLPYG DEPGYIEVD NSRIYHWDGA  
 101 GAEPDQDVL LQPFDDNGN YITDEYSAA HPTWITRLDL KEVQDVITSE  
 151 INFQSEPPVS LPVTRSPA PVAPAPAPAP MPIRGLM

!!AA\_SEQUENCE 1.0  
 P1:S76937 - hypothetical protein - *Synechocystis* sp. (strain PCC 6803)  
 C:Species: *Synechocystis* sp.  
 A:Variate: PCC 6803  
 C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 21-Aug-1998  
 C:Accession: S76937  
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hikosawa, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.; DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.  
 A:Reference number: S74322; MUID:97061201  
 A:Accession: S76937  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1312 <RAN>  
 A:Cross-references: EMBL:D90917; GB:AB001339; NID:g1653836; PID:d1019582; PID:g1653939  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 C:Genetics:  
 A:Start codon: GTG

S76937 Length: 312 February 11, 2000 15:52 Type: P Check: 9556

1 MIMKILITGG GGYIGSVLTP TLLAGYQVT VIDNPFQON SLAECQYET  
 51 FNVIRGDCRK EDLKEQKLT ADVIILPAL VGAPICSRQ IGTITNQTA  
 101 VEMICQLASP QQRILMPYTN SGYIGSGKRG FCTESPLRP ISLYGVTKVE  
 151 AEKAVIARQW SMTEPLAVF GMAPRMAYD LVNDVYRAF YDRAVVIIEG  
 201 HFRKRYIHIR DVAKYFLHGL ENESKAKGP YVVGLEDANL SKTELCAEIR  
 251 KILPNFVILE APIGEDPDKR DIVVSNORIL STGETPDMSL GRGIOELIKG  
 301 YTLRNSYVS NV

!!AA\_SEQUENCE 1.0  
 P1:S74351 - hypothetical protein s110071 - *Synechocystis* sp. (strain PCC 6803)  
 C:Species: *Synechocystis* sp.  
 A:Variate: PCC 6803  
 C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 21-Aug-1998  
 C:Accession: S74351  
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hikosawa, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.; DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium

Synechocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.

A:Reference number: S74322; MUID:97061201

A:Accession: S74351

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-333 <KAN>

A:Cross-references: EMBL:D64001; GB:AB001339; NID:g1001102; PID:d1010920;

PID:g1001128

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

S74351 Length: 333 February 11, 2000 15:52 Type: P Check: 4892 ..

1 MRKLYFLPG LGGRFVCGGL WAEKTIKLY QECPATIV YKORENTLE

51 LDDILSQDL GOVIEVISWG FDIPIKLAKL QGRVITHAH STGYGDLPT

101 ALPIVCISRN TMGYWGRSP HSLIYLPNH ISEDFSLGL VADIDVLQA

151 RKSHILKO LIPALSOQN VKVIDSYED LTGIFNRKV YLYDSLEYA

201 QGVSEGFGL CPENMACGC HVESSVNGI SDYLDGFCN EKIAGYALDY

251 DCRILGVYK GSRQWRVDP NLIAEYRREN ILERLVTLL DINEFDYHP

301 SFSOTIEPL QRIAMLEK SLITKYKNKE KSS

IIA\_SEQUENCE 1.0

P1:S74733 - hypothetical protein sl10931 - Synechocystis sp. (strain PCC 6803)

C:Species: Synechocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 21-Aug-1998

A:Accession: S74733

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.;

Miyajima, N.; Hirosewa, M.; Sugiyama, M.; Sasamoto, S.; Kimura, T.; Hosouchi,

T.; Matsuno, A.; Muraiki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimp, S.;

Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium

Synechocystis sp. PCC6803. II. Sequence determination of the entire genome and

assignment of potential protein-coding regions.

A:Reference number: S74322; MUID:97061201

A:Accession: S74733

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-156 <KAN>

A:Cross-references: EMBL:D90901; GB:AB001339; NID:g1651897; PID:d1017617;

PID:g1651958

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

S74733 Length: 156 February 11, 2000 15:52 Type: P Check: 376 ..

1 MASTKILCK TFIASGLAF LGMATVPMLT VLRGNANOTE GPSQGTPOOP

51 TADLEBRKE VAGYEYKIQ REDPNNALQ GLVENRIOMG DLGAIAPME

101 KIKRIPERE GLKOLELAIR LQYKNPOLPP PONPLAPTEP NNOGGTGGT

151 PTPROP

IIA\_SEQUENCE 1.0

P1:S74723 - hypothetical protein sl10939 - Synechocystis sp. (strain PCC 6803)

C:Species: Synechocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 21-Aug-1998

A:Accession: S74723

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.;

Miyajima, N.; Hirosewa, M.; Sugiyama, M.; Sasamoto, S.; Kimura, T.; Hosouchi,

T.; Matsuno, A.; Muraiki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimp, S.;

Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium

Synechocystis sp. PCC6803. II. Sequence determination of the entire genome and

assignment of potential protein-coding regions.

A:Reference number: S74322; MUID:97061201

A:Accession: S74723

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-125 <KAN>

A:Cross-references: EMBL:D90901; GB:AB001339; NID:g1651897; PID:d1017607;

PID:g1651948

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

S74723 Length: 125 February 11, 2000 15:52 Type: P Check: 8425 ..

1 MELIDHLYGN LVSTIRFVLE SVSYLCILG LVKTLQWVA GDRRTMTPK

51 FENSRILQFG LMLALAEFO LGADILSTTI APTLESKRL GLIAVIRTEL

101 NYFLGKELES EIRMEIEQER QEHGR

IIA\_SEQUENCE 1.0

P1:S75160 - hypothetical protein sl11635 - Synechocystis sp. (strain PCC 6803)

C:Species: Synechocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 21-Aug-1998

A:Accession: S75160

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.;

Miyajima, N.; Hirosewa, M.; Sugiyama, M.; Sasamoto, S.; Kimura, T.; Hosouchi,

T.; Matsuno, A.; Muraiki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimp, S.;

Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium

Synechocystis sp. PCC6803. II. Sequence determination of the entire genome and

assignment of potential protein-coding regions.

A:Reference number: S74322; MUID:97061201

A:Accession: S75160

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-215 <KAN>

A:Cross-references: EMBL:D90903; GB:AB001339; NID:g1652127; PID:d1017607;

PID:g1652150

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

S75160 Length: 215 February 11, 2000 15:52 Type: P Check: 4206 ..

1 MDVRFISLTK PELVIOGEPL SPEGLIAYCA RVSSPNOENP NYTKLOFCI

51 REGHSITREM VDMTELETTI RAIAPQILRH RSFSFOEFL RYSCATEYEC

101 YEARRQDVKN RQNSLDDEDE STKKWFNOAO AAWEKSHOL YEELAKGIA

151 KECARSILPL NTVTRLYMG SVRSWIRHS VRCDQATQKE HREIALARK

201 IFMKHEPTVA AALEW

IIA\_SEQUENCE 1.0

P1:S77220 - hypothetical protein slr1442 - Synechocystis sp. (strain PCC 6803)

C:Species: Synechocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 21-Aug-1998

A:Accession: S77220

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.;

Miyajima, N.; Hirosewa, M.; Sugiyama, M.; Sasamoto, S.; Kimura, T.; Hosouchi,

T.; Matsuno, A.; Muraiki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimp, S.;

Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium

Synechocystis sp. PCC6803. II. Sequence determination of the entire genome and

assignment of potential protein-coding regions.

A:Reference number: S74322; MUID:97061201

A:Accession: S77220

Mon Feb 14 08:07:32 2000

pir.cat

Page 140

A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1285 <KAN>  
A:Cross-references: EMBL:D90907; GB:AB001339; NID:g1652618; PID:d1018287;  
PID:g1652634  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

S77220 Length: 285 February 11, 2000 15:52 Type: P Check: 3780

1 MLTVHRDARD YQIFLCSEFL SLGILARDWS LSPVIAVYL ATALYOSTLF  
51 TINOQGNFSR WPSLDTWOSL WRHOGKRSAL ITGLGCLLL RANHLSTLIL  
101 AASLAIAKRF LLOAWGKRMF NPANFGIIAA LTLTGDAWVS PGOGITLML  
151 GAVFLAAGGL VLKVGKRWDT SIMFLAVYAG LPLARNAMWG WPEVSTFHL  
201 ENGSLVFLAL FMLTDFRSIP NARSSRLIWA TAIATFSLIL QYHYFLPTAL  
251 FVALFCLSPV TVLDRQFPA PRFQWQKKN PVPSN

!!AA\_SEQUENCE 1.0  
P1:S77428 - hypothetical protein ser2060 - *Synechocystis* sp. (strain PCC 6803)  
C:Species: *Synechocystis* sp.  
A:Variety: PCC 6803  
C>Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 21-Aug-1998  
C:Accession: S77428  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.  
A:Reference number: S74322; MUID:97061201  
A:Accession: S77428  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 179 <KAN>  
A:Cross-references: EMBL:D90906; GB:AB001339; NID:g1652492; PID:d1018264; PID:g1652610  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C:Genetics:  
A:Start codon: GTG

S77428 Length: 79 February 11, 2000 15:52 Type: P Check: 4795

1 MVLGNNLP TKIGAKLTMD SGRIVAVTG AIIILAVAY LVVQIIDFR  
51 GENVPAPMET MGLLTHITAI ANGLSLPFS

!!AA\_SEQUENCE 1.0  
P1:B41838 - Vans - *Enterococcus faecium*  
C:Species: *Enterococcus faecium*  
C>Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 18-Nov-1994  
C:Accession: B41838  
R:Arthur, M.; Molinas, C.; Courvalin, P.  
J. Bacteriol. 174, 2582-2591, 1992  
A:Title: The Vans-Vant two-component regulatory system controls synthesis of desipptide peptidoglycan precursors in *Enterococcus faecium* B4147.  
A:Reference number: A41838; MUID:92210502  
A:Contents: B4147; PIP816  
A:Accession: B41838  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1384 <KAT>  
A:Note: sequence extracted from NCBI backbone (NCBI:94048, NCBI:94052)

B41838 Length: 384 February 11, 2000 15:52 Type: P Check: 3600

1 LVTKLNKRN DYSKLERLY MYIAIVVA IYEVLYISM IRGLGCMIT  
51 SIENKYNLDN HLDAMKLYOY SIKNNIDIFI YVAIYISILI LCRMWSKFA  
101 KYDEINTGI DVLQNEQDQ IELSAEMDVW ECKLTKART LERREDDAKL  
151 AEOKNKVVA YLADIKPTL TSLIGLSL DEAPMPVDQ KAYVHTLTD  
201 KAYRELQID EFFETIRNL QITLTKTHI DLYIKLVOMT DEFFPOLSAH  
251 GKQAVIHAE DLTVSGDPK LARFVNIUK NAAVSEDNS IDITAGLSG  
301 DVVSIEFKNT GSIPIKDLAA IFEKFRLEN ARSSDTGAG LGLAIARETI  
351 VQGGQIYAE SNNYITFRV ELPAEDLVD KRRS

!!AA\_SEQUENCE 1.0  
P1:F69937 - conserved hypothetical protein ypj6 - *Bacillus subtilis*  
C:Species: *Bacillus subtilis*  
C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 24-Sep-1998  
C:Accession: F69937  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo, J.; Bertero, M.G.; Bessières, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Buisson, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conteron, I.F.; Cummings, N.J.; Daniel, R.A.; Denzot, F.; Devine, K.M.; Duesterhoelt, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Follmer, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Gallizzi, G.; Galleon, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Goldlight, E.C.; Grandi, A.; Gutseppl, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holstappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Kjaerz-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.  
S:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medina, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogilava, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetalle, D.; Porwilk, S.; Prescott, A.M.; Prescan, E.; Pujic, P.; Punelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.  
A:Authors: Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.U.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terstra, P.; Tognoni, A.; Totsu, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vasseroth, A.; Viat, A.; Wandut, R.; Wedler, E.; Wedler, H.; Weitzeneget, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A:Reference number: A69580; MUID:98044033  
A:Accession: F69937  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1224 <KON>  
A:Cross-references: GB:299115; GB:AL009126; NID:92634478; PID:e1183692; PID:92634665  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: ypj6

F69937 Length: 224 February 11, 2000 15:52 Type: P Check: 3316

1 MYNADYVAFG ASDDVEIGM GGTAKFVKQ EKKVMCDLT EALSSNGTV  
51 SLRKEAALA ARIIGADKRI QLTIPDRGLI MSDQAIRSVI TVIRICPKKA  
101 VFAPYKDRH PDHGNAALV EALIFSAGH KYKDEKSLPA HKYSKYYVM  
151 INGFQPDFV IDISDTIAK KSLINAKSO FTSPKDSVST PLTNGYIEIV



## 201 EAREKLYGKE AGVEAYVSEF PNGC

!!AA\_SEQUENCE 1.0  
PI:539739 - efflux protein homolog ywf - Bacillus subtilis  
N:Alternate names: protein lpa 84d  
C:Species: Bacillus subtilis  
C:Date: 07-Oct-1994 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999  
A:Accession: S39739; 670053  
R:Glasner, P.; Kunst, F.; Arnaud, M.; Coudart, M.P.; Gonzalez, W.; Hullo, M.F.; Ionescu, M.; Lubochinsky, B.; Marcelino, L.; Moszer, I.; Presecan, E.; Santana, M.; Schneider, E.; Schweizer, J.; Vertes, A.; Rapoport, G.; Danchin, A.  
Mol. Microbiol. 10, 371-384, 1993  
A:Title: Bacillus subtilis genome project: cloning and sequencing of the 97 kb region from 325 degrees to 333 degrees  
A:Reference numbers: S39655; M01D:95020537  
A:Accession: S39739; nucleic acid sequence not shown; translation not shown  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-394 <GLA>  
A:Cross-references: EMBL:X7124; NID:9413923; PID:9414008  
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1993  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessestres, P.; Bolydin, A.; Borchert, S.; Boriss, R.; Boursier, C.V.; Boursier, A.; Braun, M.; Bridgell, S.C.; Bron, S.; Brouillet, S.; Bursch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conerton, J.P.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabre, J.; Fariol, E.  
Nature 350, 249-256, 1997  
A:Authors: Roullet, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Guin, S.Y.; Glasner, P.; Goffeau, A.; Golightly, E.J.; Grandi, A.; Guiseppl, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karmali, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koettler, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Landin, S.  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, A.; Muehl, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Presecan, E.; Pulic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serrot, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Tompita, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenberg, M.; Vannier, P.; Vassartoli, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzengesser, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis  
A:Reference number: A69580; M01D:98044033  
A:Accession: G60055  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-394 <KUN>  
A:Cross-references: GB:299123; GB:AL009126; NID:92636240; PID:el186270; PID:92636306  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: ywf  
C:Keywords: transmembrane protein  
S39739 Length: 394 February 11, 2000 15:52 Type: P Check: 3644

1 KROLPNRSKY LTYGALSFM GDYCVLPALL ILSTYHYDYV VTSQVYVRS  
51 IPMVPOPLG VLVDRDLRIK IMWTIDING IIFGLTFLP KSEYPLIFLA

101 LEFTYSGV FENPARLAW SLESIDKSI NULFAATTI SIYGAAG  
151 ILLGGSVEL AVAFNGVTL VSAFISRIK LQVFPQSEN IKAQSFKE  
201 GKREIKNSF VLNAMFTMT MALMGVYS YFPIVSRFLG DQIGNFLI  
251 FGIGFGFIFG AALYKXWGN NNRGITFTV LSIYSLATL FPIPAVSVI  
301 AALFFIME YGEVLAKEYV OENANQOG RITSVAEASL GUCISGSMF  
351 INILSAPIM GLIYVYICL FHTLVKVS FLERDNTEQ KQVE  
!!AA\_SEQUENCE 1.0  
PI:42365 - flagellar assembly protein fliH - Bacillus subtilis  
N:Alternate names: fliH protein homolog  
C:Species: Bacillus subtilis  
C:Date: 24-Jul-1992 #sequence\_revision 24-Jul-1992 #text\_change 24-Sep-1998  
A:Accession: C42365; D69624; S14436  
R:Albertini, A.M.; Caramori, I.; Crabb, W.D.; Scoffone, F.; Galizzi, A.  
J. Bacteriol. 173, 3573-3576, 1991  
A:Title: The fliH locus of Bacillus subtilis is part of a large operon coding for flagellar structures, motility functions, and an Arpase-like polypeptide.  
A:Reference number: A42365; M01D:91258343  
A:Accession: C42365  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-208 <ALB>  
A:Cross-references: EMBL:X56049; NID:939904; PID:939906  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessestres, P.; Bolydin, A.; Borchert, S.; Boriss, R.; Boursier, C.V.; Boursier, A.; Braun, M.; Bridgell, S.C.; Bron, S.; Brouillet, S.; Bursch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conerton, J.P.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabre, J.; Fariol, E.  
Nature 390, 249-256, 1997  
A:Authors: Roullet, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Guin, S.Y.; Glasner, P.; Goffeau, A.; Golightly, E.J.; Grandi, A.; Guiseppl, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karmali, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koettler, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Landin, S.  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, A.; Muehl, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Presecan, E.; Pulic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serrot, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Tompita, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenberg, M.; Vannier, P.; Vassartoli, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzengesser, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis  
A:Reference number: A69580; M01D:98044033  
A:Accession: D69624  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-208 <KUN>  
A:Cross-references: GB:299112; GB:AL009126; NID:92633902; PID:el185214; PID:92633995  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: fliH  
C:Keywords: fliH  
C42365 Length: 208 February 11, 2000 15:52 Type: P Check: 1150

1 MARVEEADR ISEANSHE NRRQIDOEK NDMAEKOKL IEAKAGEFE  
 51 OGVALGRAEA MKOYALIEQ ANITEMSRK AVEKDKEDAN EEIYELVAL  
 101 AKRWQKSD DKEAEELLQO QVINEKVED DISIYDPY YETIFQKDE  
 151 IOOLLYECCR LGIYADEKAQ KGTGYETPF GRVDSVDTQ LMOCKDLT  
 201 ALENGAEE

!!AA-SEQUENCE 1.0  
 P1:069820 - hypothetical protein yhaY - Bacillus subtilis  
 C/Species: Bacillus subtilis  
 C/Date: 05-Dec-1997 #sequence-revision 05-Dec-1997 #text-change 24-Sep-1998  
 C/Accession: C69820  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berto, M.G.; Bessières, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brigelli, S.C.; Bron, S.; Brulliet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.R.; Codani, J.J.; Connetton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabre, C.; Ferrati, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandt, G.; Gutseppl, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaer-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kunita, K.; Lapidus, A.; Lardinois, S.  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Mostl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Pressecan, E.; Fujic, P.; Purnelle, B.; Rapoport, G.; Ray, M.; Reynolds, T.; Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.  
 A:Authors: Schleib, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serr, S.J.; Serr, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.; Tepstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vanlier, F.; Vasseroelli, A.; Viati, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
 A:Reference number: A69580; M01D:98044033  
 A:Accession: C69820  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-83 <KUN>  
 A:Cross-references: GB:299109; GB:AL009126; NID:92633260; PID:e1182983; PID:92633317  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: yhaY

C69820 Length: 83 February 11, 2000 15:52 Type: P Check: 5172 ..

1 MTSLEQOP FIAKAKWMS RLYWLVDR1 RSKOPFSLIS LFRFTKNHVS  
 51 GTPCLTMDAL RFLTCLCLSP LYNGMIDRI HUK

!!AA-SEQUENCE 1.0  
 P1:E69976 - hypothetical protein yrkG - Bacillus subtilis  
 C/Species: Bacillus subtilis  
 C/Date: 05-Dec-1997 #sequence-revision 05-Dec-1997 #text-change 24-Sep-1998  
 C/Accession: E69976  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berto, M.G.; Bessières, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brigelli, S.C.; Bron, S.; Brulliet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.R.; Codani, J.J.; Connetton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabre, C.; Ferrati, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandt, G.; Gutseppl, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaer-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kunita, K.; Lapidus, A.; Lardinois, S.  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Mostl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Pressecan, E.; Fujic, P.; Purnelle, B.; Rapoport, G.; Ray, M.; Reynolds, T.; Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.  
 A:Authors: Schleib, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serr, S.J.; Serr, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.; Tepstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vanlier, F.; Vasseroelli, A.; Viati, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
 A:Reference number: A69580; M01D:98044033  
 A:Accession: E69976  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-36 <KUN>  
 A:Cross-references: GB:299117; GB:AL009126; NID:92634966; PID:e1183881; PID:2635097  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: yrkG

E69976 Length: 36 February 11, 2000 15:52 Type: P Check: 1724 ..

1 MTKVMPFKE YTKSLTKNL YLFMVCWAKM ILVIGR

!!AA-SEQUENCE 1.0  
 P1:S39656 - lipoteichoic acid biosynthesis protein dlte - Bacillus subtilis  
 C/Species: Bacillus subtilis  
 C/Date: 07-Oct-1994 #sequence-revision 26-May-1995 #text-change 17-Mar-1999  
 C/Accession: S39656; G69616  
 R:Glaser, P.; Kunst, F.; Arnaut, M.; Coudart, M.P.; Gonzalez, W.; Hullo, M.F.; Ionescu, M.; Lubochinsky, B.; Marcelino, L.; Moszer, I.; Pressecan, E.; Santana, M.; Schneider, E.; Schweizer, J.; Verites, A.; Rapoport, G.; Danchin, A.  
 M01: Microbiol. 10, 371-384, 1993  
 A>Title: Bacillus subtilis genome project: cloning and sequencing of the 97 kb region from 325 degrees to 333 degrees  
 A:Reference number: S39655; M01D:95020537  
 A:Accession: S39656  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-252 <GLA>  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1993  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berto, M.G.; Bessières, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brigelli, S.C.; Bron, S.; Brulliet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.R.; Codani, J.J.; Connetton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabre, C.; Ferrati, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandt, G.; Gutseppl, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaer-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kunita, K.; Lapidus, A.; Lardinois, S.  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Mostl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Pressecan, E.; Fujic, P.; Purnelle, B.; Rapoport, G.; Ray, M.; Reynolds, T.; Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.  
 A:Authors: Schleib, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serr, S.J.; Serr, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.; Tepstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vanlier, F.; Vasseroelli, A.; Viati, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
 A:Reference number: A69580; M01D:98044033  
 A:Accession: E69976  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-36 <KUN>  
 A:Cross-references: GB:299117; GB:AL009126; NID:92634966; PID:e1183881; PID:2635097  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: yrkG

Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kashara, Y.; Kjaer-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kunita, K.; Lapidus, A.; Lardinois, S.

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Poll, T.M.; Portetelle, D.; Porwollik, S.; Prescott, A.M.; Prescann, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Ray, M.; Reynolds, T.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiuchi, J.; Sekowska, A.; Serror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Taccioni, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A:Reference number: A69580; MUID:98044033

A:Accession: G69610

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-252 <KUN>

A:Cross-references: GB:299123; GB:AL009126; NID:g2636240; PID:el166353; PID:g2636389

A:Experimental source: strain 168

C:Genetics:

A:Gene: *dlte*

C:Superfamily: short-chain alcohol dehydrogenase homology

F:6-183/Domain: short-chain alcohol dehydrogenase homology <SADH>

S39656 Length: 252 February 11, 2000 15:52 Type: P Check: 2631

1 MKMTNTVLI TGSAGIGLE LAKRLLELN EVIIGRSEA RLAEAKOQLP

51 NHTKQCDVA DSCREALYE WAKEXPNIN VLNNAGIOL EIDFKKTEE

101 LFDGDEIEL NQAPVHLA LFTPHLMKOP EAAIVQVTS LAFNPNAVY

151 VYCATRAALH SFTSLRHHV RDTSEVIEY APPNVDGILN QKSRDKGLT

201 YRGSSSEYV QYFDLGLKEG KQETNERVE GLRDATRADY DRLFQKNTQ

251 EN

!!AA\_SEQUENCE 1.0

P1:572921 - hypothetical protein B2168\_CL172 - *Mycobacterium leprae*

C:Species: *Mycobacterium leprae*

C>Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Sep-1997

C:Accession: S72921

R:Smith, D.R.; Robison, K.

Submitted to the EMBL Data Library, November 1993

A:Description: *Mycobacterium leprae* cosmid B2168.

A:Reference number: S72580

A:Accession: S72921

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-86 <SMI>

A:Cross-references: EMBL:U00018; NID:g467037; PID:g467073

S72921 Length: 86 February 11, 2000 15:52 Type: P Check: 2352

1 MTSNGPSAR DSAGKSVRDT GSFEGQPRT QFLTVAEVAL IMRVSKMTYV

51 RLVHNGELPA VRVGRSFRVH AKAVHDMLET SYFDVG

!!AA\_SEQUENCE 1.0

P1:D0669 - coenzyme F420-dependent N5, N10-methylenetetrahydromethanopterin dehydrogenase homolog - *Mycobacterium tuberculosis* (strain H37RV)

N:Alternate names: hypothetical protein Rv2951c; hypothetical protein u0002p

C:Species: *Mycobacterium tuberculosis*

C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 21-Nov-1998

C:Accession: D70669; S73070

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; 537-544, 1998

Nature 393, 537-544, 1998

A:Authors: Sgares, R.; Sultson, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence.

A:Reference number: A70500; MUID:98295987

A:Accession: D70669

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-381 <COL>

A:Cross-references: GB:Z83018; GB:AL123456; NID:g3261671; PID:e283386; PID:g1694883

A:Experimental source: strain H37RV

R:Smith, D.R.; Robison, K.

Submitted to the EMBL Data Library, September 1994

A:Description: *Mycobacterium tuberculosis* cosmid thc2.

A:Reference number: S73053

A:Accession: S73070

A:Molecule type: DNA

A:Residues: 1-49, 194, 197-381 <SMI>

A:Cross-references: EMBL:U00024; NID:g560506; PID:g560512

C:Genetics:

A:Gene: Rv2951c

D70669 Length: 381 February 11, 2000 15:52 Type: P Check: 7905

1 MGLRFGFVD ALVHSRLPPT LPARSSMAA TWAGADSYW GDHNLALVPR

51 SIATSEVLGI AKFVPKIDA NYEPMTWLN LAFGLPSRLR LGVCTADAGR

101 RNPATVQAQ ALHLHLTRGR ALDIGYGR EGNEPYGVEM TKPARPEEA

151 LATRALMNS NGELISRESP YFPLHNLFD LPPYRKWME IWVAHGRPM

201 LRATGRRADA WPIPIVYRPS DYSRALFAVR SAASDGRNP MSITPAVRG

251 IITGRNDV EEALESVYK MTAIGVPEA WARGVHEPM GADSGVODI

301 IPOTMDKQTV LSYAKVRPA LKEVVFSGT PDEVIDVAE WRDGLRIVV

351 LINGSLVNS LRKTVTAIAP HAKVLRGLK L

!!AA\_SEQUENCE 1.0

P1:C70828 - hypothetical protein Rv0463 - *Mycobacterium tuberculosis* (strain H37RV)

C:Species: *Mycobacterium tuberculosis*

C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998

C:Accession: C70828

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; 537-544, 1998

Nature 393, 537-544, 1998

A:Authors: Sgares, R.; Sultson, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence.

A:Reference number: A70500; MUID:98295987

A:Accession: C70828

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-97 <COL>

A:Cross-references: GB:AL021933; GB:AL123456; NID:g3261529; PID:e1252533;  
PID:g2909539  
A:Experimental source: strain H37RV  
A:Gene: RV0463  
C70828 Length: 97 February 11, 2000 15:52 Type: P Check: 2874 ..

1 MTRRASTDP QIINGAIGV VTEGYIMLAA ISVGGGLTIV SQMSRYVLLL  
51 SYLVAVCGAA GGLRLRSNGK LAWSAFASFSL PIPPVLTIVA VLADIYL

!!AA\_SEQUENCE 1.0  
P1:D70510 - hypothetical protein RV0519c - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998  
C:Accession: D70510  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltham, D.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Kirogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
N:ature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.  
A:Reference number: A70500; MUID:98295987  
A:Accession: D70510  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-300 <COL>  
A:Cross-references: GB:297831; GB:AL123456; NID:g3261825; PID:e329668;  
PID:g2826722  
A:Experimental source: strain H37RV  
A:Gene: RV0519c  
D70510 Length: 300 February 11, 2000 15:52 Type: P Check: 2571 ..

1 MURRCAGNT DRGIMTPMA DLTRALLRW GAGAGAGAAG VMAFGALVDP  
51 LEPQAPAPF EPTAGSSLP TRISGFISA ARGGIKTNV ISMPQSQSQ  
101 LRPVIALHGR DGNAGMLDL GVEOGLARLV KEGRPAPAVV GVDGNTYWH  
151 RSSGGDSGA MYLDELPLML TSMGMDTSRV GFLGSMGGY GALLGARLG  
201 PARTAGICAI SPALFTSFTG STPGAFDSYD DYQHSYVGL PALNSIFLRV  
251 DCGTSDRFYF ATRQFVNLH QPPAGSFSPG GHDSYVMEQ LPELMAWMS

!!AA\_SEQUENCE 1.0  
P1:F70547 - hypothetical protein RV0547c - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 12-Feb-1999  
C:Accession: F70547  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltham, D.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Kirogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
N:ature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.  
A:Reference number: A70500; MUID:98295987  
A:Accession: F70547

A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-294 <COL>  
A:Cross-references: GB:295558; GB:AL123456; NID:g3261781; PID:e316667;  
PID:g2114009  
A:Experimental source: strain H37RV  
A:Gene: RV0547c  
C:Superfamily: short-chain alcohol dehydrogenase homology  
F:41-223/Domain: short-chain alcohol dehydrogenase homology <SADH>  
F70547 Length: 294 February 11, 2000 15:52 Type: P Check: 3126 ..

1 MSKRLRWL EQLTAGMRP PISPOLLNR PAMQVPLDTG KRILLTGASS  
51 GIGAAATKCP GLHRAVVAV ARKDLIDAV ADRITGDSGT AMSLPCLSD  
101 MEALDALVED VEKRIIGDI LINNAGSR RPLASLERW HDVERTVNL  
151 YVAPRLIRG LAPGLERGD GHIINVAWG VISEASPLFS VYNSKRAALS  
201 AVSRILTEW GSGGVHSTLL YPLVATPMI APTKAYDGLP ALTPAEAEW  
251 MVTARTRPV RIAPRVAVAV NALDSIGPRW VNALMQRNE QLNP

!!AA\_SEQUENCE 1.0  
P1:C70708 - hypothetical protein RV0776c - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998  
C:Accession: C70708  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltham, D.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Kirogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
N:ature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.  
A:Reference number: A70500; MUID:98295987  
A:Accession: C70708  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-259 <COL>  
A:Cross-references: GB:280226; GB:AL123456; NID:g3261638; PID:e26566;  
PID:g1550654  
A:Experimental source: strain H37RV  
A:Gene: RV0776c  
C70708 Length: 259 February 11, 2000 15:52 Type: P Check: 4949 ..

1 MYEVGDILAM AGRNPTGYAA VDAEDCLVY GAARDASVL AALRPVYGD  
51 CLVAFDAPLV VANRTGQRA EALNDRFQ FQAGAYPANT EKPEFADVPR  
101 AARLAROLAI DMDPLSSATR RAIEVYPHRA TYALFRLPRA LTKYAKDGRS  
151 VDLKSELLE LMDGVGIAQ AGVAMOVAQ PDWVSLRQV TVAQRKSDLR  
201 AAEDPIDAVV CAYVALYQR RPADVTIYGD FTTGYIVTPS LPTDFRTAPD  
251 AGRARARAR

!!AA\_SEQUENCE 1.0  
P1:G70716 - hypothetical protein RV0953c - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998  
A:Accession: G70716

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A: Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A: Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence.  
 A: Reference number: A70500; MUID: 98295987  
 A: Accession: G70716  
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A: Molecule type: DNA  
 A: Residues: 1-282 <COL>  
 A: Cross-references: GB:Z79700; GB:AL123456; NID:93261628; PID:e264100;  
 PID:91524209  
 A: Experimental source: strain H37Rv  
 A: Genetics:  
 A: Gene: RV0953c

G70716 Length: 282 February 11, 2000 15:52 Type: P Check: 883 ..  
 1 MHYGLVLETS DRGITPAAA FLAESHGERT FYYPEHTHP VKRQAHHTT  
 51 GDASLPDDRY MRLDPWVSL GAASAVTSRI RLATAVALV EHDPTLAKS  
 101 IATLDHLSHG RVSVGVFGM NDELVDHGV PGRRTMTLR EYLEMRLM  
 151 TOEACDYDGE FVKFGPSWAM PKVPVPHIV LVGAAGTEKN EKWLRASDG  
 201 WITPRVDVI DEPVKLLODI WAAAGRDGLP QIVALDVKPV PDKLARMAEL  
 251 GVTVELFGMP DRASDDAAY VERLAKLAC CV

!!AA\_SEQUENCE 1.0  
 P1:D70897 - hypothetical protein RV1102c - *Mycobacterium tuberculosis* (strain H37Rv)  
 C: Species: *Mycobacterium tuberculosis*  
 C: Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998  
 C: Accession: D70897  
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A: Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A: Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence.  
 A: Reference number: A70500; MUID: 98295987  
 A: Accession: D70897  
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A: Molecule type: DNA  
 A: Residues: 1-103 <COL>  
 A: Cross-references: GB:AL021997; GB:AL123456; NID:93260022; PID:e1251981;  
 PID:94896739  
 A: Experimental source: strain H37Rv  
 A: Genetics:  
 A: Gene: RV1102c

D70897 Length: 103 February 11, 2000 15:52 Type: P Check: 9641 ..  
 1 MRPHIAOLD KAPVILIR EVVPHLTV TVAPITTVR GLATEVADA  
 51 VNGINPSV SCNDNTQITV CDLGRQIGYL LASQEPALAE AIGNAFDLWD  
 101 VVA  
 !!AA\_SEQUENCE 1.0

P1:F70958 - hypothetical protein RV1377c - *Mycobacterium tuberculosis* (strain H37Rv)  
 C: Species: *Mycobacterium tuberculosis*  
 C: Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998  
 C: Accession: F70958  
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A: Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A: Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence.  
 A: Reference number: A70500; MUID: 98295987  
 A: Accession: F70958  
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A: Molecule type: DNA  
 A: Residues: 1-212 <COL>  
 A: Cross-references: GB:Z81011; GB:AL123456; NID:93242274; PID:e275148;  
 PID:91621259  
 A: Experimental source: strain H37Rv  
 A: Genetics:  
 A: Gene: RV1377c

F70958 Length: 212 February 11, 2000 15:52 Type: P Check: 4095 ..  
 1 MPGIDFALY RGSPSEGLP PITPPWDTK APRDNVIGMH TGGVHGDVL  
 51 DIGCGLDGNA IYLRNGQYV TGLDISPTAL TTAKRRASDA GVDVFAVGD  
 101 ATKRLGYTGA FDTVIDGMP HCLDDDGKRS YAASVHRATR PGATLLSCF  
 151 SNAMPDEEM PRSTVSEQL RDLVGAGWD IESLEPAIVR RELDGEVEM  
 201 AFMNVRAQR GS

!!AA\_SEQUENCE 1.0  
 P1:C70763 - hypothetical protein RV1558 - *Mycobacterium tuberculosis* (strain H37Rv)  
 C: Species: *Mycobacterium tuberculosis*  
 C: Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998  
 C: Accession: C70763  
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A: Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A: Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence.  
 A: Reference number: A70500; MUID: 98295987  
 A: Accession: C70763  
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A: Molecule type: DNA  
 A: Residues: 1-148 <COL>  
 A: Cross-references: GB:Z74020; GB:AL123456; NID:93261584; PID:e248539;  
 PID:91403484  
 A: Experimental source: strain H37Rv  
 A: Genetics:  
 A: Gene: RV1558

C70763 Length: 148 February 11, 2000 15:52 Type: P Check: 497 ..  
 1 MPLSGEAPS PLDMSEQAD TYMSGGTEG TOLGKFPVIL LTVGATGK  
 51 LRKTPMRVE HDGQVAIVAS LGAPKNPVW YHWVKNPRV ELDDGTVTGD

101 YDAREVGEDE KAIMWORA VMPDYASYQT KTDROIPEV LPEVRAGG

11AA\_SEQUENCE 1.0  
P1:G70541 - hypothetical protein Rv1577c - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998  
C:Accession: C70541

R:CoLe, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence

A:Reference number: A70500; MUID:98295987

A:Accession: C70541  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-170 <COL>

A:Cross-references: GB:295586; GB:AL123456; NID:93261785; PID:e317180;  
A:Experimental source: strain H37Rv

C:Genetics: Rv1577c

C70541 Length: 170 February 11, 2000 15:52 Type: P Check: 1907 ..

1 MAERISEGR TVHGTIVPN EATVRDPDG EFOEMFPGA FRISIERGH

51 KILLVSHDA RTRYPVGRV ELREPHGLF GAFELDTPD GDEALAVKA

101 GVVDSESVGF RPIRDRREGD VLVRVEALL EVSLGVPRV SGAQIAVRA

151 ESTLVVRSRT AEAWLSILDW

11AA\_SEQUENCE 1.0  
P1:A70639 - hypothetical protein Rv1957 - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998  
C:Accession: A70639

R:CoLe, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence

A:Reference number: A70500; MUID:98295987

A:Accession: A70639

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-181 <COL>

A:Cross-references: GB:284498; GB:AL123456; NID:93261701; PID:e293662;  
A:Experimental source: strain H37Rv

C:Genetics: Rv1957

A70639 Length: 181 February 11, 2000 15:52 Type: P Check: 9160 ..

1 MTRDRDADL DLQVGAFLA AQAQRDRRL LRQAAVHRA PRPAQLIYD

51 LEFEAVADAD PATISAFVR ISCHLRIONQ AADDVKEGD TKDETQVAT

101 ADEFAALFD YHLOEGEDD TEELTAYAA TTGRFALYPR IREYVDLTG

151 RLALPELIE ILSRPMVSP GAQWPAIRGT P

11AA\_SEQUENCE 1.0  
P1:B70758 - hypothetical protein Rv1996 - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998  
C:Accession: B70758

R:CoLe, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence

A:Reference number: A70500; MUID:98295987

A:Accession: B70758

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-317 <COL>

A:Cross-references: GB:274025; GB:AL123456; NID:93261586; PID:e248804;  
A:Experimental source: strain H37Rv

C:Genetics: Rv1996

B70758 Length: 317 February 11, 2000 15:52 Type: P Check: 1836 ..

1 MSAQOINLGI VVGVDSPCS HTAVWAARD AQMRVALRV VQVPPVITA

51 PEGNAEYRS FOEAKREIV EHSYLAQAH QIVEQAKVA LEASSCGRA

101 QITGEVLHQ IYPTLANIS QVAMVLGYR GQAVAGALL GSVSSTLVH

151 AHGPAVAVE EPRPARPPA PVVVGIDGSP TSGLAELAF DESRGGVDL

201 VALHANSKDG PLDFRLNWA PIEMRLNDE QEKMLARLS QMODRPDVP

251 VHRVVCDBR APRLELAQT AOLVVGSHG RGFPGHNG SVSRVAVNSG

301 QAVVIVARIP QDPVAPA

11AA\_SEQUENCE 1.0  
P1:A70941 - hypothetical protein Rv2018 - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 26-Aug-1999  
C:Accession: A70941

R:CoLe, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence

A:Reference number: A70500; MUID:98295987

A:Accession: A70941

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-239 <COL>

A:Cross-references: GB:AL021899; GB:AL123456; NID:93242282; PID:e1251997;  
A:Experimental source: strain H37Rv

C:Genetics: Rv1996

PID:92896755

Mon Feb 14 08:07:32 2000

pir.cat

Page 147

A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: RV2018

A70941 Length: 239 February 11, 2000 15:52 Type: P Check: 7719 ..

1 MAGDELELH EDVPLYLAE ASRYLVPRRA TLATWDGZE RRPANAPVQ  
51 GGPITLALPH PTGSHARLPE VGIAEAYVIN AFRRAGVPMQ RIRPSLMLI  
101 KXVGPALALAS QDLCTDGAEV LWFPAERSGE GSPDDLVVRG LIVPRSGQIV  
151 FREIVHYIQ QISFADNLA SMIRLPQYGD ANVILDPERG YGQPVFDSSG  
201 VRYADVGLPL RAGATFQAVA DDYGVTPDOL RDAIDAIVA

!!AA\_SEQUENCE 1.0  
P1:D70943 - hypothetical protein RV2037c - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998  
C:Accession: D70943  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagsels, K.; Krogh, A.; Mclean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.  
A:Reference number: A70500; PMID:98295987  
A:Accession: D70943

A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-324 <COL>  
A:Cross-references: GB:AL021899; GB:AL123456; NID:g3242282; PID:e1252016; PID:q2896774  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: RV2037c

D70943 Length: 324 February 11, 2000 15:52 Type: P Check: 7125 ..

1 MALVSTARVD LYCEGGGVRG IGLVGVADAL ADAGYRPRV AGSSAGAIYA  
51 SLVAALQTAG EPVRLAEM RSIDYKFLD RNLIGHVPLI GGGLSLILSD  
101 GYVRGAVLBD LLGGLLADLG VHTFGDLRTG EAPDFAMSL VTAASLSRR  
151 RLVRIPDDLD SYGHPDDFS VARAVHASSA IPFVEPVAV RGATWVDSGL  
201 LSNFVALFD RTDAEPNMT FGILSRAPG IPRTPVQGV VSLGIAIET  
251 LVSNQNDAYI DDPCTVARI FVPAHVSPI DEDITAEORE ALYORGFQAG  
301 OKFIANNVYA DCLADCGPF TPST

!!AA\_SEQUENCE 1.0  
P1:H70863 - hypothetical protein RV2449c - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998  
C:Accession: H70863  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagsels, K.; Krogh, A.; Mclean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.  
A:Reference number: A70500; PMID:98295987  
A:Accession: H70863

A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-419 <COL>  
A:Cross-references: GB:AL021246; GB:AL123456; NID:g3261507; PID:e1237574; PID:g2791489  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: RV2449c

H70863 Length: 419 February 11, 2000 15:52 Type: P Check: 4861 ..

1 MTATPREFDI VLXGATFVG KLAIEYLARA GGDARIALAG RSTGRVLAVR  
51 EALGESAOVW PILPADASLP STLQAMARA QVYVTVGVY TRYGLPLVAA  
101 CAAAGTDVAD LTGEPMERN SIDLYHKOA DTGARIVHAC GFDSVPDLS  
151 VVALYHARE DGAGELDTN CVRSFKGE SGGTISMLE VLASTANDPD  
201 ARROLSDPYM LSPDRGAPE LQPODLPSS RGRRLAPELA GWNAGFLMA  
251 PNTTRIVRS NALIDWAYGR RFRYSRTMSV GSTVLAPVVS VYGGVGNAM  
301 FGASVYIHL LPRGIVKRV PRGTGPSMA AREGVYRIE TYTTTTCAR  
351 YIARMOQDD PGYKATSVIL GEGGLALALD RDLKLDWGV LTPAAMGDA  
401 LIRLTPAGV SIQTRLAS

!!AA\_SEQUENCE 1.0  
P1:E70572 - hypothetical protein RV2622 - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998  
C:Accession: E70572  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagsels, K.; Krogh, A.; Mclean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.  
A:Reference number: A70500; PMID:98295987

A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-273 <COL>  
A:Cross-references: GB:295387; GB:AL123456; NID:g3261763; PID:e315941; PID:q2104289  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: RV2622

E70572 Length: 273 February 11, 2000 15:52 Type: P Check: 8454 ..

1 MANKRGAG PLESDRDD HMOGHWLLAR LGRVLRPG VELTRTLAR  
51 AEVTDADVLE LAPGIGRTAA EILARNRSY VGEASDPNAA NLVRYVLAR  
101 GDVATDAD TGLSDASDV VIGEAINTQ GNAKHITVA EAARVLRPG  
151 RYAIHELAV PDVAQVORT DLRSIARAL KYNARPLVA EWSHLIAGHG

```
201 LVEHVHTAS MALLOPRRI ADEGLGALR FAGNLIHRA ARRRVLLMRH
251 TFRHRERLT AVAIVAKPH VDS

!!AA_SEQUENCE 1.0
P1:F70572 - hypothetical protein Rv2623 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
C:Accession: F70572
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeyer, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
A:Reference number: A70500; MUID:98295987
A:Accession: F70572
A>Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-297 <COL>
A:Cross-references: GB:295387; GB:AL123456; NID:g3261763; PID:g316002;
PID:g2104288
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv2623

F70572 Length: 297 February 11, 2000 15:52 Type: P Check: 9524 ..

1 MSSGSSLSGI IVGIDSPAA QVAVRMAARD AELRKIPPLT VHAVSEVAT
51 WLEVPILPPGV LRMQODGRH LIDDALKVE QASIRAGPPT VHSIIVPAAA
101 VPTLIDMSKD AVLNVVGLG SGRWPGRLG SVSSGLRHA HCPVVIHDE
151 DSVMPHPOA PVLGVGDSS ASELATIAIF DEASRRANDL VALHANSVDV
201 VSEWPGIDWP ATGSMAGQVL AERLAGMOR YPNVAIRRV VROGPARQVL
251 QRSEEAQLVY VGRGRGQYA GMLVGSVGET VAQLARPPVI VARERLT

!!AA_SEQUENCE 1.0
P1:G70965 - hypothetical protein Rv2651c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
C:Accession: G70965
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeyer, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
A:Reference number: A70500; MUID:98295987
A:Accession: G70965
A>Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-177 <COL>
A:Cross-references: GB:280225; GB:AL123456; NID:g3242265; PID:g266401;
PID:g1550692
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv2651c

G70965 Length: 177 February 11, 2000 15:52 Type: P Check: 4149 ..

1 MSSILFRTAE LRPGEGRTY GVIVPYGEVT TVRDLDEGR EMPAFGARFR
51 SIAMRGKRV LVSMDARTR YPGRAVELR EEPHGIFGAF ELANTPGDE
101 ALANVAKGV DAFSVGRPI RDRREGDVI RVEALALEVS LTGPALGA
151 QIAGVRAESL AVVSSSLAED RIALMDV

!!AA_SEQUENCE 1.0
P1:E70885 - hypothetical protein Rv2859c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
C:Accession: E70885
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeyer, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
A:Reference number: A70500; MUID:98295987
A:Accession: E70885
A>Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-308 <COL>
A:Cross-references: GB:AL008883; GB:AL123456; NID:g3261490; PID:e1172951;
PID:g2612804
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv2859c

E70885 Length: 308 February 11, 2000 15:52 Type: P Check: 9649 ..

1 MDLSASRSDG GDLRPASPR LRSPVSDGCD PLRASPRLR SPVSDGGDPL
51 RPASPRLRSP LGASRPVGL TAYLEQVRTG VMDIRAGIYP ADYREGITMA
101 GGVAVLLPPQ PYDPESVGCY LDSLHALVIT GGYDDPPAY GQEPHPRATDH
151 PRGRDAMEF ALLRGALORG MPVLIGICGT QVYNALGCT LHOHLPDILG
201 HSGHRAQNGV FTRLPVHTAS GTRLAELIGE SADVCYHHQ AIDOVGEGLY
251 VSAVDVGVV EALRLPGDF VLAQWHPER SDDLRLFRA LVDAASGYAG
301 RQSAEPR

!!AA_SEQUENCE 1.0
P1:F70699 - hypothetical protein Rv2953 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 05-Dec-1998
C:Accession: F70699; S73068; S73067
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeyer, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
```



A:Reference number: A70500; MUID:98295987  
 A:Accession: F70669  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-418 <COL>  
 A:Cross-references: GB:283018; GB:AL123456; NID:93261671; PID:e283385;  
 PID:g1694681  
 A:Experimental source: strain H37RV  
 A:Submitted to: EMBL Data Library, September 1994  
 A:Description: Mycobacterium tuberculosis cosmid tb02  
 A:Reference number: 573053  
 A:Accession: 573058  
 A:Molecule type: DNA  
 A:Residues: 1-171, 'RAAFIATLIGRLGNSVPSNAHGIGRR', 'GPSARHPVMDHGPGRRTTCAAGFSASRT'  
 <SMIL>  
 A:Cross-references: EMBL:U00024; NID:9560506; PID:9560514  
 A:Accession: 573067  
 A:Molecule type: DNA  
 A:Residues: 'MNCVPTF', 172-418 <SMIL2>  
 A:Cross-references: EMBL:U00024; NID:9560506; PID:9560515  
 A:Genetics:  
 A:Gene: RV2953  
 F70669 Length: 418 February 11, 2000 15:52 Type: P Check: 5693 ..  
 1 MSPAREEPDI VLYGATGFSG KITAHELAHS GSTARIALAG RSSERLGVR  
 51 MMLGPAAMW PLILDASOP LILEMAARA QVVLTVGPY TRYGLPLVAA  
 101 CAAAGDLYAD LTGELMFCRN SIDLYHROA DTGARIIILAC GPDSPSDLN  
 151 VYGLYRSYE DGTGELCTD LVLRSPQRM VSGGSVATYS EAMRTASDP  
 201 EARLYTDPY TLTTRGAEP ELGAPDPELR RPRGLAPEL AGFTGGEVQ  
 251 APENTRYVR SNAIDENAYG RRRYSSTMS LKSNAAPI LAAVTGIVAG  
 301 TIGLKNKYD RLPRLYERV TPKRGIPSR KTOERGHTE EYTTTTTGA  
 351 RYRATFANV DAYKSTAVIL AQSGLALAD RDLRLNELGV LTPAAAGDA  
 401 LLARLPAGV VMGTRLS  
 1:IAA-SEQUENCE 1.0  
 P1:F70651 - hypothetical protein RV3073c - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998  
 C:Accession: B70651  
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeyer, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagers, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: B70651  
 A:Molecule type: DNA  
 A:Residues: 1-118 <COL>  
 A:Cross-references: GB:283867; GB:AL123456; NID:93261691; PID:e290998;  
 PID:g1781140  
 A:Experimental source: strain H37RV  
 A:Genetics:  
 A:Gene: RV3073c

B70651 Length: 118 February 11, 2000 15:52 Type: P Check: 3649 ..  
 1 MVRTRVRA RYEDIDPD GORLYDRIW PGRKRDOR VGWCKDVAP  
 51 SKELREYHH OPERFDEFAS RYOEHLHDSA ALAEKLITLG RSVVTPYAT  
 101 RHVARSHAV LAQLNGR  
 1:IAA-SEQUENCE 1.0  
 P1:D70645 - hypothetical protein RV3131 - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998  
 C:Accession: D70645  
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeyer, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagers, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: D70645  
 A:Molecule type: DNA  
 A:Residues: 1-332 <COL>  
 A:Cross-references: GB:283867; GB:AL123456; NID:93261695; PID:e290954;  
 PID:g1781236  
 A:Experimental source: strain H37RV  
 A:Genetics:  
 A:Gene: RV3131  
 D70645 Length: 332 February 11, 2000 15:52 Type: P Check: 2556 ..  
 1 MNHFRPAET VRTVTLAVR APSIHNTQPM RMWVCPSTLE LFSRPMQMR  
 51 STDPGRLEI LSCGVALHC VALASLGWC AKVNRFPDPK DRCHLITIG  
 101 QPLVPDQADY ALAAPIRRR TDRAYSQWP VGGDIALMA ARAARGVML  
 151 ROVSALDRMK AYAQAVLDH VTDEYLREL TWSRGYSV AGYARNEPP  
 201 SDSPAPIGR LRAGPLSOP SDVLPADGA ATALGTEND DRLARLAGE  
 251 AASIVLTAT AMGACCPIT EPLEIATRD AVRAVEFGG GYPMMLRVG  
 301 WAPINADLP PTPRELQV VEPPELLRQ RC  
 1:IAA-SEQUENCE 1.0  
 P1:F70951 - hypothetical protein RV3200c - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998  
 C:Accession: F70951  
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeyer, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagers, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: F70951  
 A:Molecule type: DNA

A:Residues: 1-355 <COL>  
A:Cross-references: GB:AL021646; GB:AL123456; NID:g3242278; PID:e1248791;  
PID:g2827610  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: RV3200C  
F70951 Length: 355 February 11, 2000 15:52 Type: P Check: 8567 ..

1 MAGSWRLRG LNEKLTAPG VALVGLRIP QRASPARVI SRRVAVVA  
51 LLTLAGIYV DRDGLDAG DRITFLDCLY YAAVTLSTG YGDITPISF  
101 ARAINIFVIT PLRIAFILIL VGTLEVLTE TSROAKTOR MRSRVNHTV  
151 VIGYTKGKT AVAMVSDCL VGEIYVVDI DSGVLERAAA AGLVYHGD  
201 TKSVDLRLAG TQHASIIVA TSRDPAVILV TLTRRLAPK AKIVASIRA  
251 ENQHLLRQSG ADIVVVSSET AGRLLGIATF TPSVEMIED LLTPEAGLAV  
301 AEREVEQAEV GGSFRLRDI VLGVVRDGL LRIAGPEVDA IEASDRLLXI  
351 ROVGR

!!AA\_SEQUENCE 1.0  
P1:C70561 - hypothetical protein RV3626c - Mycobacterium tuberculosis (strain H37Rv)  
C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998  
C:Accession: C70561  
R:Coile, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence  
A:Reference number: A70500; MUID:98295987  
A:Accession: C70561  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-350 <COL>  
A:Cross-references: GB:295436; GB:AL123456; NID:g3261770; PID:e316560;  
PID:g2105068  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: RV3626c  
C70561 Length: 350 February 11, 2000 15:52 Type: P Check: 48 ..

1 MTGASELTIG NTVDWERFAS VGERLARPAF PSTETTRQV IDELYIAAK  
51 ABPPVADVTG LIADGVVPPA RVVDRPAMIR SAASMRAMT HGSAPRGFL  
101 TGRITGAOTG AVLAFAVSGI LQGYDPFGAA GEGCLLVYP NVIAVEROLR  
151 VEPSDFRLMV CLHEVTHRVQ ETANPLISG MSQALNLITE EPVDDIGRVV  
201 SRLADFISR GHGTDSEVN PGGILGLVRA VQSPQRKAL DQLLVIGTIL  
251 EGHASHVDA VGRWVPSVA TIRRRPDDR HHKRPPIQRL VALLGFDAK  
301 LQYTRGAF VDAVDRAG KLPTIWGSP ETLPLPAEIE NPORWIDRVL

!!AA\_SEQUENCE 1.0  
P1:H70850 - hypothetical protein RV3912 - Mycobacterium tuberculosis (strain H37Rv)  
C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998  
C:Accession: H70850  
R:Coile, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence  
A:Reference number: A70500; MUID:98295987  
A:Accession: H70850  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-254 <COL>  
A:Cross-references: GB:AL021426; GB:AL123456; NID:g3261511; PID:e1245970;  
PID:g2808697  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: RV3912  
H70850 Length: 254 February 11, 2000 15:52 Type: P Check: 8984 ..

1 MSADDPDK HSADAPPLT VELLADLQAG LLDATARI RSRVSDPQA  
51 QQLRLNRY RRDVAAGAD PANGPARPA VDSISALR SARPNSSPA  
101 AHARHVRH VMTAGAGL CAVATAGVG AVNDAPPAP SATTQHIT  
151 VSRPAPYPL SRPQVLDLH HTDYGPPG PLGDSRRTS CLSGLGTPAS  
201 TPVLAGPID IDARPVILV IPADTPDKLA VFAVAPHCSA ADTGLASTV  
251 VPRA

!!AA\_SEQUENCE 1.0  
P1:H70865 - probable cutinase precursor - Mycobacterium tuberculosis (strain H37Rv)  
C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998  
C:Accession: A70565  
R:Coile, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence  
A:Reference number: A70500; MUID:98295987  
A:Accession: A70565  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-226 <COL>  
A:Cross-references: GB:295390; GB:AL123456; NID:g3261766; PID:e316057;  
PID:g2104375  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: RV3452  
A70565 Length: 226 February 11, 2000 15:52 Type: P Check: 5027 ..

1 MIPRPHSG RWRACARRL TSLVAAFAA ATLITPALA PRASACPPA  
51 EVFAROTGE PGLGRVGOA FVSLRQTN KSIGTYGVN PANGDELAAA  
101 DGANDASDH QOMASACRAT RLVLGYSOG AAVIDIYTA PLDGLFTOP

151 LPPAADHIA AALFENPG RAGLMSALT POGSKTINL CNNGPICSD  
201 GNRMRHLGY VPGMTNOAR FVASRI

!!AA\_SEQUENCE 1.0  
P1:D70696 - probable dtdp-glucose 4 - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 16-Dec-1998  
C:Accession: D70696  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagsels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.  
A:Reference number: A70500; MUID:98295987  
A:Accession: D70696  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-326 <COL>  
A:Cross-references: GB:280343; GB:AL123456; NID:g3261648; PID:e267402;  
PID:g1552867  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: epib  
A:Superfamily: UDPglucose 4-epimerase homology  
F:3-31/Domain: UDPglucose 4-epimerase homology <UDP>  
D70696 Length: 326 February 11, 2000 15:52 Type: P Check: 8388 ..

1 MEIVTSGAG FQGSHTESL LANKHWTVL DKSSNAYRN MGRSRMDRA  
51 AFISGSVTDG QTDRAVRDH HVYHLLAAHV NNDQSLGDE SFLETNWGT  
101 YRVLEAVRRY RNELIVSTC EYVGDGHNK EGERLDENAF LKPSPIYGAS  
151 KAADRILCY S YFRSYGLDVT TVRPNTIFV RQNAAGFGAL IPRLVROGIN  
201 GEGLTIFGAG SATRDYLYVS DIVAGYNLV RPTLRGQAI NPAAGKOTRV  
251 RDIYEVADK FGARIEHRDA RRGVORFPA DISLANSIG OPOVEIMDGI  
301 DRYIMAKDQ PQPYEDDGF SGGSYVL

!!AA\_SEQUENCE 1.0  
P1:B70986 - probable isomerase - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998  
C:Accession: B70986  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagsels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.  
A:Reference number: A70500; MUID:98295987  
A:Accession: B70986  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-203 <COL>  
A:Cross-references: GB:295890; GB:AL123456; NID:g3242245; PID:e318986;  
PID:g2131010

A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: RV1745C

B70986 Length: 203 February 11, 2000 15:52 Type: P Check: 9024 ..

1 MTRSYRAPP IERYVLNDR GDATGVADKA TVHTGDTPLH LAESYVFDL  
51 HDQILLTRRA ATRKTPANW TNSCCGHPD GSLPGAIRL RLAEGLTLP  
101 DRVDLIFGE RIRAMADGT VENICEPYR VQVDDQPRN SDEVDALRWL  
151 SWEDFVRDVT AGVIAPVSW CRSOLGYLTK LGPCPQWNV ADDCRLPKAA  
201 HGN

!!AA\_SEQUENCE 1.0  
P1:H70760 - probable lipoprotein signal peptidase - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998  
C:Accession: H70760  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagsels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.  
A:Reference number: A70500; MUID:98295987  
A:Accession: H70760  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-202 <COL>  
A:Cross-references: GB:274020; GB:AL123456; NID:g3261584; PID:e248552;  
PID:g1403503  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: lspa

H70760 Length: 202 February 11, 2000 15:52 Type: P Check: 2319 ..

1 MPDEPTGSAD PLTSTBEAG AGEPNAPAP RLRLMLLSVA VVVLTDIVT  
51 KYAVVQLLPP GQPVSLIGDT VTWLVNRNG AAFSNATGTV WVLVLATGV  
101 VVGIFWNGRR LVSPWNLGL GMILGAMGN LVDRFRFAG PLRHWVDFL  
151 SVGMWPEVNV ADPSVVGAI LVLITSLIFG DEDTVGRRA DGDIVGRKA  
201 DG

!!AA\_SEQUENCE 1.0  
P1:B70763 - probable membrane protein - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998  
C:Accession: B70763  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagsels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.

A:Reference number: A70500; MUID:98295987  
 A:Accession: B70763  
 A:Status: Preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-397 <COL>  
 A:Cross-references: GB:274020; GB:AL123456; NID:93261584; PID:e248540;  
 PID:91403485  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: mmpL6

B70763 Length: 397 February 11, 2000 15:52 Type: P Check: 4418 ..  
 1 MGSISVGLV KRGWVRVVE DTIDGIDQLG EQLASTVTL DKLAIQPOL  
 51 VALIPDEIAS QQINRELALA NYATMSGIYA QTRALLENNA AMGAFDPAK  
 101 NDSFYLPPE AFDPDPFQNG LKPLSDSK AARMTISHEG DPATPEGISH  
 151 IDAIKQAAHE AVKGTPMAGA GIVLAGTAT FMDIDGATY DLTAGTIAL  
 201 SLILIMMII TRSIVAAVVI VGIYALSGLA SEGLSVLVWQ HLGIDLYWI  
 251 VLALAVILL AVGSDYNILL ISRKEEIGA GLNTGIIRAM AGTGVVTTAA  
 301 GLVFAATMSS FVFSDLRYVG QIGTTIGLGL LEDLLVRAF MPTSIAYLLG  
 351 RMEWMPQVRA PRPASRMALRP YGPRPVVREL LIREGNDPR TOYATIR

!!A:SEQUENCE 1.0  
 P1:A70614 - Probable mma2 protein - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998  
 A:Accession: A70614  
 R:Colo. S.T.; Broesh, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Besham, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Feltham, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Kirogi, A.; McLean, J.; Moule, S.; Hamlin, N.; Holtroyd, S.; Hornsby, T.; Jagers, K.; M.A. Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: A70614  
 A:Status: Preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-287 <COL>  
 A:Cross-references: GB:292772; GB:AL123456; NID:93261722; PID:e306575;  
 PID:91877382  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: mmaA2

A70614 Length: 287 February 11, 2000 15:52 Type: P Check: 5098 ..  
 1 MYNDLTPHE DYQAHYDSD DFFRLDPT QTYSCAHER EDMTEBAQI  
 51 AKIDIALGKL GLQPGMTLLD ICGMGKATIR RAIQOYDVN VGLTSSKNOA  
 101 AHYQSPDEM DTPRDRVVL AGWEQNEPV DRIVSISAFE HFCHDRHADP  
 151 FARAHKILPP DGVLLHRTIT GLTRQOMVDH GLPTLMLAR ELKFTATEIF  
 201 PGQOPTIEM VEDOSAKTGF TLTRRSLOSP HYATRLDMA EALOEKRSKA  
 251 IAIQSEYVE RYMKYITGCA KLFRRGYIDV NOFTLAK

!!A:SEQUENCE 1.0  
 P1:S1338 - Hypothetical protein (insertion sequence IS112) - Streptomyces albus  
 C:Species: Streptomyces albus

C:Date: 19-Mar-1997 #sequence\_revision 10-Oct-1997 #text\_change 30-Jun-1998  
 A:Accession: S1338  
 R:Rodicio, M.R.; Alvarez, M.A.; Chater, K.F.  
 M.Ol. Gen. Genet. 225, 142-147, 1991  
 A:Title: Isolation and genetic structure of IS112, an insertion sequence responsible for the inactivation of the SalI restriction-modification system of Streptomyces albus G.  
 A:Reference number: S1338; MUID:91155926  
 A:Accession: S1338  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-256 <ROD>  
 A:Cross-references: EMBL:X56644; NID:946594; PID:9581565  
 A:Experimental source: Insertion sequence IS112  
 A:Note: the authors translated the initiation codon GNG for residue 1 as Val  
 C:Genetics:  
 A:Mobile element: Insertion sequence IS112  
 A:Start codon: GNG

S1338 Length: 256 February 11, 2000 15:52 Type: P Check: 4079 ..  
 1 MAGVTASEP SWIAPFSGLS PROFGKLVTV LRRGADAVR KGRPMSLPLE  
 51 DRALLVAIYV RTNLTMRQLA PLEGVSKSNA DRIDHLGPM LALOPRRFA  
 101 KDTVLIVDGT LVPTRDHTIA ERSKNRYST NHQVVIDADT RLIVVVGHPRL  
 151 AGNRNDCKAM EESGAKAAG KTLITADGY PGTVGVIPHR RERGGAGLPD  
 201 WKEHNKSHK QVRAVEHVE AMKTKILR DCRILKGQVH HAMGIAMNH  
 251 NIALTG

!!A:SEQUENCE 1.0  
 P1:PC4110 - transcription regulator homolog - Streptomyces aureofaciens (fragment)  
 N:Alternate names: hypothetical 137 protein  
 C:Species: Streptomyces aureofaciens  
 C:Date: 10-Jan-1996 #sequence\_revision 08-Feb-1996 #text\_change 09-Sep-1997  
 A:Accession: PC4110  
 R:Kotmanec, J.; Lempelova, A.; Farkasovsky, M.; Homeroova, D.  
 Gene 165, 77-80, 1995  
 A:Title: Cloning, sequencing and expression in Escherichia coli of a streptomyces aureofaciens gene encoding glyceroldehyde-3-phosphate dehydrogenase.  
 A:Reference number: JCA373; MUID:96084956  
 A:Accession: PC4110  
 A:Molecule type: DNA  
 A:Residues: 1-137 <ROD>  
 A:Cross-references: GB:U21191; NID:9706951; PID:9706952  
 C:Comment: This protein has a helix-turn-helix DNA-binding domain in the carboxyl-terminus.  
 C:Keywords: transcription

PC4110 Length: 137 February 11, 2000 15:52 Type: P Check: 2351 ..  
 1 NHARLVAA PYRSGQAQY VPRSVPEVHG ERFATREMA LRLSEPLTL  
 51 ELARHAAYS ARTFSRRPAE DTGYTPMOW MRARVDARE LERSRSVE  
 101 QIADVGLGT GSNRLHPOR ILGTPSEYR RTFTGE

!!A:SEQUENCE 1.0  
 P1:PN0644 - Hypothetical protein 66 - Streptomyces coelicolor (fragment)  
 C:Species: Streptomyces coelicolor  
 C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 07-May-1999  
 A:Accession: PN0644  
 R:May, J.T.; L.V.; Fisher, S.H.  
 Gene 130, 145-150, 1993  
 A:Title: The Streptomyces coelicolor glr gene encodes a protein similar to other bacterial response regulators.  
 A:Reference number: JN0831; MUID:93345814  
 A:Accession: PN0644

Mon Feb 14 08:07:32 2000

plr.cat

Page 153

A:Molecule type: DNA  
A:Residues: 1-66 <WRA>  
A:Cross-references: GB:L03213  
C:Genetics:  
A:Start codon: GTG  
PN0644 Length: 66 February 11, 2000 15:52 Type: P Check: 9954  
1 MAKVTRDDVA RLAGSTAVV SYVINGPRP VAPATREYV LAKKEGYRP  
51 DRVAQMASR RTDLIG  
11AA\_SEQUENCE 1.0  
P1:S4423 - strp protein - Streptomyces glaucescens  
C:Species: Streptomyces glaucescens  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 16-Dec-1998  
C:Accession: S4423  
R:Mayer, G.; Piepersberg, W.  
Submitted to the EMBL Data Library, April 1994  
A:Reference number: S44224  
A:Accession: S4423  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-358 <MAX>  
A:Cross-references: EMBL:X78974; NID:g475235; PID:g581674  
C:Genetics:  
A:Start codon: GTG  
A:Superfamily: UDPglucose 4-epimerase homology  
F:3-310/Domain: UDPglucose 4-epimerase homology <UDP>  
S4423 Length: 358 February 11, 2000 15:52 Type: P Check: 6765  
1 MRLITGOG YLGTVAAPV LTAHQVYGL DSGLFASRVL GALVDVYVG  
51 LALDRDVTY DTLRGYDAV HLAALSNPL GSDPELTHA VNHASTRLA  
101 RLAKAGVTR FAYASTCSY GAGSDAPVD EDAPLKYVP YAVSKYVED  
151 DLVELADSGE APFLRNATA FGSPRLSD IVLNILGRA YLGGRTYLS  
201 DGTAMPPLVH AEDTARAVLA GLAAPATYR ARAINIGTE NNRVAEIAE  
251 AAAAAYGVS VEISGETND PSRYVDYTR ARTELGHRAT WIIPOGAQL  
301 AREYRARGLT RAFEHDFTR LAVLERRRE GSLDGLDVR PARVRDGDH  
351 RTATAEAG  
11AA\_SEQUENCE 1.0  
P1:S17674 - serine-type D-Ala-D-Ala carboxypeptidase (EC 3.4.16.4) precursor - Streptomyces sp. (strain K15)  
N:Alternate names: D-alanyl-D-alanine carboxypeptidase; DD-peptidase; penicillin-binding protein  
C:Species: Streptomyces sp.  
C:Date: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 13-Sep-1998  
C:Accession: S17674  
R:Palomque-Messia, P.; Englebert, S.; Leyh-Bouille, M.; Nguyen-Disteche, M.; Bouchet, C.; Houba, S.; Dideberg, O.; van Beumen, J.; Ghuyssen, J.M.  
Bouchet, C.; 229-230, 1991  
A:Title: Amino acid sequence of the penicillin-binding protein / DD-peptidase of Streptomyces K15. Predicted secondary structures of the low-M(r) penicillin-binding proteins of class A.  
A:Reference number: S17674; MID:92028812  
A:Accession: S17674  
A:Molecule type: DNA  
A:Residues: 1-291 <PAL>  
A:Cross-references: EMBL:X59965; NID:g47150; PID:g581684  
C:Genetics:  
A:Start codon: GTG  
C:Keywords: hydrolase; serine carboxypeptidase  
F:1-29/Domain: signal sequence #status predicted <SIG>  
F:30-291/Product: D-alanyl-D-alanine carboxypeptidase #status predicted <MAT>

S17674 Length: 291 February 11, 2000 15:52 Type: P Check: 4817  
1 MLRLRAATV ITTGALLAG TLGATPATAV TKPTIAVGS YAMNGTGT  
51 LTKKADIRR STGSTTKIM AKVVALQSNL NLDKATIQK AYSDDVANK  
101 PSQAHLYVD KVTYROLLG LMLPSGDDA YALADKSGS SOAARVRSF  
151 IGMNTAATN LGLNTHFDS FDSIGGANY STPHRLTKIA SSAMNSTFR  
201 TVVKTKAYTA KVTYKTSIR TMDTKNTNG LLSYSQAIG VTKGSGPEAK  
251 YCLVFATRG KTVIAGTVLA STSPARESD ATKIMNGFA L  
11AA\_SEQUENCE 1.0  
P1:S1717 - RNA methylase - Streptomyces tenebrarius  
N:Alternate names: gentamicin resistance protein  
C:Species: Streptomyces tenebrarius  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 20-Mar-1998  
C:Accession: S1717  
R:Holmes, D.J.; Cundliffe, E.  
Mol. Gen. Genet. 229, 229-237, 1991  
A:Title: Analysis of a ribosomal RNA methylase gene from Streptomyces tenebrarius which confers resistance to gentamicin.  
A:Reference number: S1717; MID:92017656  
A:Accession: S1717  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-280 <HOL>  
A:Cross-references: GB:S60108; NID:g237470; PID:g237471  
C:Keywords: antibiotic resistance  
S1717 Length: 280 February 11, 2000 15:52 Type: P Check: 5969  
1 MHPRAPGPD PEDPRLAEVY DAVRSSRRYQ SVAPETVRRL ATSLAVASRG  
51 DLAEAVKRTK RGLHEIFGAY LSPSPKYDAL LROLRGAVDA ATPRCGHPA  
101 PRHYARLHP RALPIIDEFY REVFARCAP ASYVDIACGW NPLAPMWP  
151 SDAFTYHSD IDTRLMEFLD AALETLGVAH DVAEYDLMTG VGEVETDYL  
201 LKTYPCIEA QGRQGWDLI DAIRSPLYVY SFPTKSIGOR SKGMENYSA  
251 NEDAWLENRP HDVQLEFRN ELVYFVRKNA  
11AA\_SEQUENCE 1.0  
P1:C64251 - hypothetical protein homolog MG464 - Mycoplasma genitalium (SGC3)  
C:Species: Mycoplasma genitalium  
C:Date: 10-Nov-1995 #sequence\_revision 10-Nov-1995 #text\_change 10-Oct-1997  
C:Accession: C64251  
R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.; Bult, C.J.; Kellaway, A.R.; Sutton, G.; Kelley, J.M.; Fritchman, J.L.; Weidman, J.F.; Small, K.V.; Sandusky, M.; Merrick, J.M.; Tomb, N.; Dougherty, B.A.; Bolit, K.F.; Hu, P.C.; Lincier, T.S.; Peterson, S.N.; Smith, H.O.; Hutchison III, C.A.; Venter, J.C.  
Science 270, 397-403, 1995  
A:Title: The minimal gene complement of Mycoplasma genitalium.  
A:Reference number: A64200; MID:96026346  
A:Accession: C64251  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-385 <TRGR>  
A:Cross-references: GB:U39733; GB:L43967; NID:g1046177; PID:g1046184; TIGR:MG464  
A:Experimental source: strain G-37  
C:Genetics:  
A:Genetic code: SGC3  
C64251 Length: 385 February 11, 2000 15:52 Type: P Check: 2836  
1 MKIKLAQNK EITTFNPFM SAAVNEKNN WKNFKRSAL FIKYIKVIF

51 IFLITVLGNG CTOTLAQPMW GTNOVLGSL EIGYKEGTG DRYDLISNN  
 101 FGPFITSDY TLAYGFYGM FWPMAQIVL PIMYATRPVL GSGVELGFNM  
 151 ILSLIVLL VRLITVITL NSTLAEKN EVOGKLAELN AKYGAIDLQ  
 201 SKNRQLEIM SLYKHNKIS SAAVQVEVT LPEFLIYRI VTLRPIKAI  
 251 ILENFMDLSK VPLTEIFSNF TTGWPFIIF LVIIVPQFL SOKLPQVMS  
 301 KRNEAKAHS OKSIEDLNKT KKMQLIEFYV FAITAFSAA GVGWYFELNA  
 351 LFTLOSXYLT HVEIVKRRK RKONYSKLD ILERE

!!AA\_SEQUENCE 1.0  
 P1:164230 - sensory rhodopsin II transducer homolog - Mycoplasma genitalium (SGC3)  
 C:Species: Mycoplasma genitalium  
 C>Date: 17-Nov-1995 #sequence\_revision 17-Nov-1995 #text\_change 10-Oct-1997  
 C:Accession: 164230  
 R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.; Bult, C.J.; Kerlavage, A.R.; Sutton, G.; Kelley, J.M.; Nguyen, D.; Uterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merriam, J.; J.F.; Dougherty, B.A.; Bott, K.F.; Hu, P.C.; Duster, T.S.; Peterson, S.N.; Smith, H.O.; Hutchison III, C.A.; Venter, J.C.  
 S:Science 270, 397-403, 1995  
 A:Title: The minimal gene complement of Mycoplasma genitalium.  
 A:Reference number: A64200; MUID:96026346  
 A:Accession: 164230  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-248 <TIGR>  
 A:Cross-references: GB:U39707; GB:U43967; NID:91045970; PID:91045975; TIGR:MG280  
 A:Experimental source: strain G-37  
 A:Genetic code: SGC3

164230 Length: 248 February 11, 2000 15:52 Type: P Check: 1048

1 MLVYFVLGFG GITALYFRST AANLYVOARN SIDSSFNASK AFANALANSA  
 51 NOEKSSITN NLDQVKKLE OSLOKVDYK KMLESONNNG NISOEKIREL  
 101 DAIKKDLNS KTOLDNFKNN LDKNGTASS PSVKKATAD GYLSAVSEIS  
 151 TOAOSIVSSY EKIKNNIPSS EGFNNYDYV MTTIVAVSGG MLAILITITV  
 201 FSEITSKRKG LIRFSRFIST EOLADHVNDI LDRYPDLEE VITALDHD

!!AA\_SEQUENCE 1.0  
 P1:573488 - hypothetical protein K05\_orf385 - Mycoplasma pneumoniae (ATCC 29342) (SGC3)  
 C:Species: Mycoplasma pneumoniae  
 A:Variety: ATCC 29342  
 C>Date: 27-Feb-1997 #sequence\_revision 25-Apr-1997 #text\_change 17-Jul-1998  
 C:Accession: S73488  
 R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.  
 S:Nucleic Acids Res. 24, 4420-4449, 1996  
 A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae.  
 A:Reference number: S73327; MUID:97105885  
 A:Accession: S73488  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-385 <HIM>  
 A:Cross-references: EMBL:AE000017; GB:U00089; NID:91673812; PID:91673823  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996  
 C:Genetics:  
 A:Genetic code: SGC3

S73488 Length: 385 February 11, 2000 15:52 Type: P Check: 5866

1 MFLNKKER ELKTFNPFW SAAVNEKNG LKNLKKAMGI IFKYLVKVAIF  
 51 IFLITVLGNG CTOTLAQPMW GTNOVLGSL EIGNFGTGT DRYDLQSNM  
 101 VGPYFESDY TLAYGFYGM FWPASQIVL PIMYATRPVL GSGPELGFNM  
 151 ILSLIVLL VRLITVITL NSTLAEKN EVOGKLAELN AKYGAIDLQ  
 201 SKNRQLEIM SLYKHNKIS SASVQVEVT LPEFLIYRI VTLRPIKAI  
 251 ILENFMDLSK VPLTEIFSNF TSTGTFIIF LVIIVPQFL SOKLPQVMS  
 301 KRNEAKAHS OKSIEDLNKT KKMQLIEFYV FAVITAFSAA GVGWYFELNA  
 351 LFTLOSXYLT HLFIVKRRR RLRTYSKLDQ MLERE

!!AA\_SEQUENCE 1.0  
 P1:A36149 - spiralin - Spiroplasma melliferum (SGC3) (fragment)  
 C:Species: Spiroplasma melliferum  
 C>Date: 30-Nov-1990 #sequence\_revision 13-Jan-1993 #text\_change 31-Oct-1997  
 C:Accession: A36149  
 R:Chevalier, C.; Sallard, C.; Bove, J.M.  
 S:J. Bacteriol. 172, 6090-6097, 1990  
 A:Title: Spiralin of Spiroplasma citri and Spiroplasma melliferum: amino acid sequences and putative organization in the cell membrane.  
 A:Reference number: A36149; MUID:91008990  
 A:Accession: A36149  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-237 <CHE>  
 A:Cross-references: GB:M59366; GB:M33991; NID:9152891; PID:9152892  
 C:Genetics:  
 A:Genetic code: SGC3  
 C:Keywords: Transmembrane protein

A36149 Length: 237 February 11, 2000 15:52 Type: P Check: 5138

1 SILAVGVSA VGTTSVYACN KTESNLSRV KTIAPATVA ASTPRANVRP  
 51 EIKTALFANY LKAVQGVYKT ATADQFQEV YKNSKGTALE TIDLKAGVE  
 101 VYLQITPARD KTVVIGETRY IKVILPKHGE VTKVDIDVT VTEQTVGKA  
 151 STPRAVKDE LNAVNTYATL AKAVLDATON IAPRAGASDP EITNNGAEGD  
 201 YEAKREVEVT VKAKNDSANI SQGFPRKAVY TATAPTE

!!AA\_SEQUENCE 1.0  
 P1:H69019 - conserved hypothetical protein MTH1148 - Methanobacterium thermoautotrophicum (strain Delta H)  
 C:Species: Methanobacterium thermoautotrophicum  
 C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 05-Jun-1998  
 C:Accession: H69019  
 R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakeley, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, L.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso, A.; Bush, D.; Saffer, H.; Patwell, D.; Prabhakar, S.; Modougall, S.; Shimer, G.; Goyal, A.; Petrokovski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
 S:J. Bacteriol. 179, 7135-7155, 1997  
 A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functional analysis and comparative genomics.  
 A:Reference number: A69000; MUID:98037514  
 A:Accession: H69019  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-157 <MTH>  
 A:Cross-references: GB:AE000884; GB:AE000666; NID:92622242; PID:92622253  
 A:Experimental source: strain Delta H  
 C:Genetics:  
 A:Gene: MTH1148

H69019 Length: 157 February 11, 2000 15:52 Type: P Check: 3145 ..

1 MIRATACFC THGFVAREH AVSNGYPCGY SMSVSDVVL VAGLFIPTLS

51 GIRSILKEP PEPSTATINDI KYTTEEDER VALMARSVR ELTSDLGIG

101 TTACIGCGGI AVSENREEV INSDVEADR FSGAEIILAR QRSIGRCLAE

151 LLESFLE

!!AA\_SEQUENCE 1.0

PI:D69049 - conserved hypothetical protein MTH1371 - Methanobacterium thermoautotrophicum (strain Delta H)

C:Species: Methanobacterium thermoautotrophicum

C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 05-Jun-1998

C:Accession: D69049

R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Alredede, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaltre, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso, A.; Bush, D.; Saefer, H.; Patwell, D.; Prabhakar, S.; McDougall, S.; Shimer, G.; Goyal, A.; Pietrokowski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.

J: Bacteriol. 179, 7135-7155, 1997

A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functional analysis and comparative genomics.

A:Reference number: A69000; MUID:98037514

A:Accession: D69049

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-355 <MTH>

A:Cross-references: GB:AE000900; GB:AE000666; NID:g2622478; PID:g2622480

A:Experimental source: strain Delta H

C:Genetics:

A:Gene: MTH1371

D69049 Length: 355 February 11, 2000 15:52 Type: P Check: 1234 ..

1 MNLKVAKME LRSTLSRKF LFIFLQJAV LILFLFSG FIDVLESGBA

51 LTPSLTGAE LGVIDPSGLI QGOLNPDVL IREGDAGSV LYVDNFTGIP

101 LNAIYLDYS DPRSYVRDE VEANVRSS LISEFVKSP GPDPEVREIT

151 RGEPLPVOL TRVWSTILF LPEVLFGLV VDSIVGKER KTGELIAMP

201 VRSEIILGK CLSVITVAL QIGVMILIL ASGFHMVNP L GAYFTVYASS

251 APVAGLIALI SVYARNRRA GIGITLAYII SAATLIPAL ATWAGSSGV

301 SPMTLTIKMI SGOAGLPTS FLQPSQSLYS ASYSMALQSG CSGGMNTSV

351 PDQYV

!!AA\_SEQUENCE 1.0

PI:D69065 - conserved hypothetical protein MTH1489 - Methanobacterium thermoautotrophicum (strain Delta H)

C:Species: Methanobacterium thermoautotrophicum

C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 05-Jun-1998

C:Accession: D69065

R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Alredede, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaltre, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso, A.; Bush, D.; Saefer, H.; Patwell, D.; Prabhakar, S.; McDougall, S.; Shimer, G.; Goyal, A.; Pietrokowski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.

J: Bacteriol. 179, 7135-7155, 1997

A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functional analysis and comparative genomics.

A:Reference number: A69000; MUID:98037514

A:Accession: D69065

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-249 <MTH>

A:Cross-references: GB:AE000909; GB:AE000666; NID:g2622596; PID:g2622605

A:Experimental source: strain Delta H

C:Genetics:

A:Gene: MTH1489

D69065 Length: 249 February 11, 2000 15:52 Type: P Check: 9991 ..

1 MRLGFSTLAL FMEPLENIQ KAEDDGFELV ELLCGEPYWP RRLLODGLSL

51 EYFESFDLEV LIHAPITDNL PASMNRGIRE EGRGMVETI ELMSRGAT

101 VTHRPGVVR REDRRSAL QFALFLGEC VEAADLSIK FSVENMPGRF

151 SYLCNPAEH ERFVKGCSY ATVDIGHANT TGRLODFLEI KRRAHYISD

201 NNGKROHLP LGEVTVDLKL LGSIERGVIE LNSYDGVIRS RLILEVVR

!!AA\_SEQUENCE 1.0

PI:G64362 - 2-Isopropylmalate synthase (EC 4.1.3.12) - Methanococcus jannaschii

C:Species: Methanococcus jannaschii

C>Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 10-Oct-1997

C:Accession: G64362

R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; FitzGerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kesteven, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, D.; Nguyen, D.; Utterback, T.R.; Geoghegan, N.S.M.; Waldman, J.E.; Fuhmann, J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

A:Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.

A:Reference number: A64300; MUID:96337999

A:Accession: G64362

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-406 <BUL>

A:Cross-references: GB:U67500; GB:L77117; NID:g1591202; PID:g1591206; TIGR:MJ0503; PID:g1510577

C:Genetics:

A:Map position: FOR444536-445756

C:Keywords: carbon-carbon lyase; oxo-acid-lyase

G64362 Length: 406 February 11, 2000 15:52 Type: P Check: 6767 ..

1 MTKVLYWEND FLFENSKAV CPYPRKLDK DIYIDTILR DGEOTPEVCF

51 TRECKLEIAR KIDELGLKQI EAGFPYVSR EADIYKTIN EGLNADILAL

101 CRALKNDIK AIECDYDGI TFIATSPHL KTKFNKSID EILEMGYEA

151 EYAKHGLEV AFSADART PIEDLIKVKH AAEAGADRV HIADTGCAT

201 POSNEFTICK LKRNKKAHI GVCHNDGFG AVINSTYGLI GKAAVSTTV

251 NGIGERAGNA ALEETIMALT VLYVDIGLN LEVLPCLRM VEESGILMP

301 KNRPIVEVL FKHESGIAVD AVIENPLTYE FLPEKIGLK RNILGKHS

351 CRAVAYKLK MGIDYDREL CEIVKVKET REBGEKITDE VKFEIVEVL

401 RKRKNK

!!AA\_SEQUENCE 1.0

PI:G64456 - hypothetical protein homolog MJ1251 - Methanococcus jannaschii

C:Species: Methanococcus jannaschii

C>Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 10-Oct-1997

C:Accession: G64456

R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann, J.L.; Nguyen, D.; Uitterback, T.R.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
 Science 273: 1058-1073, 1996  
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*.  
 A:Reference number: A64300; MUID:96337999  
 A:Accession: D64456  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-167 <BOL>  
 A:Cross-references: GB:067565; GB:L77117; NID:g1591874; PID:g1591884;  
 TIGR:MJ1251; PID:g1511255  
 C:Genetics:  
 A:Map position: FOR1194471-1194974

C64456 Length: 167 February 11, 2000 15:52 Type: P Check: 7004 ..

1 MITATACCF THANIGLTIH KAAAGYEDFE FKYLFSBEDL KIMKNVRVIS  
 51 AMFVPSIGV EKLDIRLPE PDFNYKYAKA YSEKDELEVA KIMAGGLAKK  
 101 LNVNISIGST AGVGRGALICI LFDNNRYLFT SDVYANLITF EIKERQKNG  
 151 IEKGRKFL EILKREYF

11AA\_SEQUENCE 1.0  
 P1:D64423 - hypothetical protein MJ0988 - *Methanococcus jannaschii*  
 C:Species: *Methanococcus jannaschii*  
 C>Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 10-Oct-1997  
 C:Accession: D64423  
 R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann, J.L.; Nguyen, D.; Uitterback, T.R.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
 Science 273: 1058-1073, 1996  
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*.  
 A:Reference number: A64300; MUID:96337999  
 A:Accession: D64423  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-329 <BOL>  
 A:Cross-references: GB:067541; GB:L77117; NID:g1591641; PID:g1591650;  
 TIGR:MJ0988; PID:g1511021  
 C:Genetics:  
 A:Map position: FOR918363-919352

D64423 Length: 329 February 11, 2000 15:52 Type: P Check: 9981 ..

1 MELLEYLRKD EYLFCHHNA DPAVAGSCVA LKYLASQNP NGKFRISADS  
 51 VSKLSRNILN EIGERVDIET YPKLPETVFI VDTASINOLK VNFDELKERE  
 101 VILIDHKKKT DLADICKYYI IKEDYPTST EIIAEIFKELN IFFPKNVRIA  
 151 LLGGIVDYTK HLKLANSTF ELISYLKDI SFQKILYLS QESDVKRKA  
 201 HLKCSRMEI REFDKLRIAL SHVSSHEASC AKTIVSIGAD VAFVAVVRKK  
 251 EKEIRVSNRC RKHVSKYVHL GNLMEKIKKE LGSGGSGHSE AGGLNAPYDK

301 SKSEKRVINE VLNLCKRFV EEXKKAON

11AA\_SEQUENCE 1.0  
 P1:D64452 - hypothetical protein MJ1221 - *Methanococcus jannaschii*  
 C:Species: *Methanococcus jannaschii*  
 C>Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 10-Oct-1997  
 C:Accession: D64452  
 R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann, J.L.; Nguyen, D.; Uitterback, T.R.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
 Science 273: 1058-1073, 1996  
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*.  
 A:Reference number: A64300; MUID:96337999  
 A:Accession: D64452  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-299 <BOL>  
 A:Cross-references: GB:067563; GB:L77117; NID:g1591850; PID:g1591852;  
 TIGR:MJ1221; PID:g1511223  
 C:Genetics:  
 A:Map position: FOR1164064-1164963  
 A:Start codon: TTG

D64452 Length: 299 February 11, 2000 15:52 Type: P Check: 2920 ..

1 MMEGEMRYM KIIPKFLN TVEILIKRNN AYSISIEPL KISIEGIIII  
 51 TCNADARDAE KIYELKKLG LGEKGHGYT IMPANTIFSC REGGISTSL  
 101 SPEELIYKAK TWKTKITNVI IKVILASING VIGLEHNIP TLIGAMITAP  
 151 LVDTVNGSAI GTVLGDELEF IOGKREKELC SGIVYACFI PSLEFVSKEL  
 201 VLQYLSSTSI ILSAIVAIIA GISGGSINAS GKEYEIIQVY IDVSLIIPAL  
 251 LMGMALATMD LYLITITFIL LAIIVILVDV GYIGLKKRY GKNIKRIKY

11AA\_SEQUENCE 1.0  
 P1:F69396 - acyl-CoA dehydrogenase, short chain-specific (acds) homolog - *Archaeoglobus fulgidus*  
 C:Species: *Archaeoglobus fulgidus*  
 C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 05-Jun-1998  
 C:Accession: F69396  
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwin, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kerlavage, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirness, E.F.; Dougherty, B.A.; McKenney, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
 Nature 390: 364-370, 1997  
 A:Authors: Uitterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujili, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus*.  
 A:Reference number: A69250; MUID:98049343  
 A:Accession: F69396  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-437 <KLE>  
 A:Cross-references: GB:AE001023; GB:AE000782; NID:g2689346; PID:g2649411;  
 TIGR:AF1175

F69396 Length: 437 February 11, 2000 15:52 Type: P Check: 9110 ..



```

1 MWMKPYLIDL DELSEEDRML KEHYRFPAEE VIRPASIELD RMPPEERKVP
51 GSPFEKWKQ IKKLGYHMY LPENKGGPGL TPLORILPE EIANGSIGFA
101 TALGVDAIPE VSAAFGTPE IYEDLVYPMI EDESGFHGC WGVTEPHGS
151 DYLITREEP ELVAKFGKN VYAEKDGEM VISGQSAWY SSAPATFHC
201 LHAOLKGRS IAEGLFCIYP LDAGGVKKR PADMLGMRD PQGELFFDGV
251 RIPEHYVVA PGFFYGFED OLLCLTSCM GAFVGLAAR CPEELHYAR
301 QRVGGVDIV KHKVKIKLY EMFEVEYELAR YYVRALEYI HKKIEQOTA
351 DASPHARAA QIYAKIAYE VAHDLQJFG AYGLIKFVI EXFFDARCM
401 LIEDGIVEVL SLEAEVDIV NTEKEFYDVN HYKAVLG

!!AA_SEQUENCE 1.0
P1:G69440 - conserved hypothetical protein AF1528 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-Jun-1998
C:Accession: G69440
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum,
K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.;
Kerlavage, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush,
J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.;
McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.;
Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.;
McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Uitterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.;
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-136 <LFE>
A:Cross-references: GB:AE000997; GB:AE000782; NID:g2689320; PID:g2649036;
TIGR:AF1528

G69440 Length: 136 February 11, 2000 15:52 Type: P Check: 1599 ..

1 MDVYALIQI LTKMKVPI MLYSLISNA PFLSPQFSRV GEFINILIRE
51 SKNSGFAIA SFLIPIVGI SLSAMIRAG KLSFRETACS AALSLMRG
101 RAALFLAPV AVSFGVEIG LFEVSLDIS RLAVFL

!!AA_SEQUENCE 1.0
P1:A69470 - hypothetical protein AF1762 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-Jun-1998
C:Accession: A69470
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum,
K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.;
Kerlavage, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush,
J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.;
McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.;
Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.;
McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Uitterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.;
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-148 <DAS>
A:Cross-references: EMBL:AF016485; NID:g2822278; PID:g2822413
A:Experimental source: strain NRC-1
A:Genome: plasmid pNRC100

A:Molecule type: DNA
A:Residues: 1-61 <KFE>
A:Cross-references: GB:AE000981; GB:AE000782; NID:g2689304; PID:g2648794;
TIGR:AF1762

A69470 Length: 61 February 11, 2000 15:52 Type: P Check: 6749 ..

1 NMPILKRDG KLAETILNRP EVHNFESWAT LTRGFAEKVY TMLKRRRNS
51 TEIKFLCRHA F

!!AA_SEQUENCE 1.0
P1:D69509 - hypothetical protein AF2077 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-Jun-1998
C:Accession: D69509
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum,
K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.;
Kerlavage, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush,
J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.;
McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.;
Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.;
McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Uitterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.;
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-270 <LFE>
A:Cross-references: GB:AE000960; GB:AE000782; NID:g2689283; PID:g2648461;
TIGR:AF2077

D69509 Length: 270 February 11, 2000 15:52 Type: P Check: 4730 ..

1 MKKLLILAA TLVVLGSSG NFRYYGARG VAVNIADNNS SYIGFECPEM
51 TLVYLANDST GVISVKNNG EDVLYESTP NDILTFSNPT FLVAGEEKLV
101 EAERGSQGE YILPIDIEAF WDNGSAKIPA CSKVAVDPAV RMSKVLDSGN
151 TVPLFTREV WKFRILVESD AGDNYTILDY IPAEFEVLSI AASDGSFATH
201 HPGAGRSCTT KIYWDYVVG TEPMDYIAT KKNPAGKQEF TSPGSYMLND
251 GAETKIGLIV SNPIYKAVR

!!AA_SEQUENCE 1.0
P1:T08352 - hypothetical protein H1337 - Halobacterium sp. (strain NRC-1)
Plasmid pNRC100
C:Species: Halobacterium sp.
A:Variety: strain NRC-1
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C:Accession: T08352
R:Ng, W.Y.; Cliffo, S.A.; Smith, T.M.; Bumgarner, R.E.; Baslin, D.; Faust, J.;
Hall, B.; Lorez, C.; Seto, J.; Slagel, J.; Hood, L.; Dassarma, S.
Genome Res. 8, 1131-1141, 1998
A:Title: Snapshot of a large dynamic replicon in a halophilic Archaeon:
megaplasmid or mitochondrion?
A:Reference number: Z16408; MUID:99063795
A:Accession: T08352
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-148 <DAS>
A:Cross-references: EMBL:AF016485; NID:g2822278; PID:g2822413
A:Experimental source: strain NRC-1
C:Genetics:
A:Genome: plasmid pNRC100

```

Mon Feb 14 08:07:32 2000

pl.r.cat

Page 158

```
T08352 Length: 148 February 11, 2000 15:52 Type: P Check: 7002
1 MSNSPEETHD FDMFPGVD GEGCIRIARA DREYSGFE LAPMLRITHE
51 QLTGLDAEG WIGVKVONS EXYVDHRLQ RIEVTEYRD HLMQALCAYC
101 DTHDVNCHVS SLEPNEDRSQ QYMGVRSIN NTKITLTPLR DQFIYKRE

!!AA:SEQUENCE 1.0
P1:S73242 - hypothetical protein 38 - red alga (porphyra purpurea) chloroplast
C:Species: chloroplast Porphyra purpurea
C>Date: 19-Mar-1997 #sequence_revision 09-May-1997 #text_change 10-Sep-1997
C:Accession: S73242
R:Reith, M.; Munholland, J.
A:Title: Complete nucleotide sequence of the Porphyra purpurea chloroplast
genome.
A:Reference number: S73108
A:Accession: S73242
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-291 <REI>
A:Cross-references: EMBL:U38804; NID:q1276653; PID:q1276787
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October
1995
C:Genetics:
A:Keywords: chloroplast

S73242 Length: 291 February 11, 2000 15:52 Type: P Check: 4681
1 MPAFAKKEE LKPIKEDTP KYTSYIIIE IEALVORLEL QVWRPATIM
51 AGIOPLEML VLEGGLPCNA PVNLFTINTS YNRELSGII VETSEFGLAN
101 SGLPMEDRE FGLNRLTLA PLISRTSIIF SSATFMTOLS LIQVIFIVA
151 SLFPMGNPLS NSSTLIFALI VLLVTVGM LSLALSTLP GHIELLALIL
201 VVNLPLEFS TALAPLYEMP PWLOLIASLN PLSYAIKIR YIYNTDMNF
251 TESVIRISWG DISLOIISL LLELDVIGAY IVSNILKARL N

!!AA:SEQUENCE 1.0
P1:S20471 - class V zygote-specific protein - Chlamydomonas reinhardtii
C:Species: Chlamydomonas reinhardtii
C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Mar-1998
C:Accession: S20471
R:Matters, G.L.; Goodenough, U.W.
Mol. Gen. Genet. 232, 81-88, 1992
A:Title: A gene/pseudogene tandem duplication encodes a cysteine-rich protein
expressed during zygote development in Chlamydomonas reinhardtii.
A:Reference number: S20471; MID:92204139
A:Accession: S20471
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-86 <MAT>
A:Cross-references: GB:S90874; NID:9247969; PID:9247970

S20471 Length: 86 February 11, 2000 15:52 Type: P Check: 4134
1 MKGSAKLTA LAILATFSA THAGPIAYGI CQTGCNALAV ACYAAAGTF
51 GYPSWGAAP ATVLACNAGL GTCMAACVAA GUSPIP

!!AA:SEQUENCE 1.0
P1:S25969 - hypothetical protein 69 - liverwort (Marchantia polymorpha)
mitochondrion
C:Species: Marchantia polymorpha
C>Date: 07-May-1993 #sequence_revision 07-May-1993 #text_change 17-Mar-1999
C:Accession: S25969
R:Oda, K.; Yamato, K.; Ohta, E.; Nakamura, Y.; Takamura, M.; Nozato, N.;
Akashi, K.; Kanehara, T.; Ogura, Y.; Kohchi, T.; Ohgama, K.

J. Mol. Biol. 223, 1-7, 1992
A:Title: Gene organization deduced from the complete sequence of liverwort
Marchantia polymorpha mitochondrial DNA. A primitive form of plant
mitochondrial genome.
A:Reference number: S25941; MID:92114051
A:Accession: S25969
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-69 <ODA>
A:Cross-references: EMBL:M68929; NID:9786182; PID:9786194
A:Note: the nucleotide sequence was submitted to the EMBL Data Library,
February 1992
C:Genetics:
A:Genome: mitochondrion
A:Keywords: mitochondrion

S25969 Length: 69 February 11, 2000 15:52 Type: P Check: 7679
1 MASLSTFRF QFLTPATF IYPAHFSHF GFAFGQLH LFGPGGLFE
51 RENAHCHIEP NKTTPRIQ

!!AA:SEQUENCE 1.0
P1:T01531 - hypothetical protein A_IG005110.4 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 24-Mar-1999
C:Accession: T01531
R:Andrews, S.
Submitted to the EMBL Data Library, July 1997
A:Description: The sequence of A. thaliana IG005110.
A:Reference number: Z14347
A:Accession: T01531
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-289 <AND>
A:Cross-references: EMBL:AF013293; NID:92252823; PID:92252827
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4
A:Insertions: 156/1
A:Note: A_IG005110.4

T01531 Length: 289 February 11, 2000 15:52 Type: P Check: 1750
1 MVAHAKSATI IPKERLKNRI VEHGRLAQR PTPLANIITY LMLPGFILS
51 IIRYFNPLP PERFARYTE MUGIHLITRG HRPSPSPGT LGNIYVYNHR
101 TALDEIYVAI ALGRKICVT YSVRLSML SPRAVALTR DRATDANMR
151 KLEKGLDVI CPBGTGCEE YLLEFSLFA ELSDRIVPA MNCKQGNFG
201 TTVGAVFMD PYFFFNPRP SYEATFDRL PEDMTVNGG KTPLEVANY
251 QKVGAVLGF ECTELTRDK YLLGNGDK VESINTRK

!!AA:SEQUENCE 1.0
P1:T05405 - hypothetical protein F10M6.130 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
C:Accession: T05405
R:Bevan, M.; Weichselgartner, M.; Farlmann, B.; Granderath, K.; Dauner, D.;
Heizl, A.; Neumann, S.; Hohsels, J.; Mewes, H.W.; Meyer, K.F.X.; Schueller, C.
submitted to the Protein Sequence Database, February 1998
A:Reference number: Z15414
A:Accession: T05405
A:Molecule type: DNA
A:Residues: 1-222 <BEV>
A:Cross-references: EMBL:AL021811
A:Experimental source: cultivar Columbia; BAC clone F10M6
C:Genetics:
A:Map position: 4
A:Insertions: 39/1, 100/1, 131/3, 172/2
```

Mon Feb 14 08:07:32 2000

pir.cat

Page 159

A:Note: F10M6.130  
T05405 Length: 222 February 11, 2000 15:52 Type: P Check: 4069 ..  
1 MTGLVSLFA SVERNREELT IWRVEODDE EHWICVCEET LICVAMELO  
51 LGRGRGVTA GRCESSLMR SKKSEGTASF AVETVLMAGF AGHFWYRGA  
101 TLERLALSTRA IDEKSKIRL LKREVEAIEL KPSNPLPIN DKAILGVGI  
151 PLHRQOYTT RAHRTIRRT ISPLMGLSHE LVRRFGSTIE IAIGFSGSRV  
201 KGLPFKEDL MKNSNEDLD LG  
!!AA\_SEQUENCE 1.0  
P1:T00559 - hypothetical protein F1216.15 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse ear cross)  
C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 30-Apr-1999  
C:Accession: T00559  
R:Bounslay, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.  
submitted to the EMBL Data Library, July 1998  
A:Description: Arabidopsis thaliana chromosome II BAC F1216 genomic sequence.  
A:Reference number: 214168  
A:Accession: T00559  
A:Status: translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-296 <ROU>  
A:Cross-references: EMBL:AC004218; NID:93355463; PID:93355478  
C:Genetics:  
A:Experimental source: cultivar Columbia  
A:Map position: 2  
A:Note: F1216.15  
T00559 Length: 296 February 11, 2000 15:52 Type: P Check: 9966 ..  
1 MEHRKDIIS CLPLELLIYI ISFLPESAR LPLVSTRFR SVNQAIIYA  
51 HSHNGSTEDI SHAVSRFINN FDEHDSKNT RTELHVDKS TFWSTIIAPH  
101 NVMHSGFFS GSKKESEFC WRLEIDQIP RRVDSGFLV KTLCLDSVNS  
151 LTRHVVSNW LEFLSDSLK ICCKRRLTSL TIDSPKILH LSISGCPKLR  
201 VLEIISFKIK TPHYGSLPL IKIHENLT KAFDVTOGP SYNNALDIG  
251 PLSLTIKNSQ SLTICRMWE VTKIVSKFEC SKIIRKIYVT IFVNRK  
!!AA\_SEQUENCE 1.0  
P1:T01617 - hypothetical protein F19F24.9 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse ear cross)  
C:Date: 19-Feb-1999 #sequence\_revision 19-Feb-1999 #text\_change 14-May-1999  
C:Accession: T01617  
R:Bounslay, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.  
submitted to the EMBL Data Library, April 1998  
A:Description: Arabidopsis thaliana chromosome II BAC F19F24 genomic sequence.  
A:Reference number: 214153  
A:Accession: T01617  
A:Status: translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-392 <ROU>  
A:Cross-references: EMBL:AC003673; NID:93004543; PID:93004564  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Map position: 2  
A:Note: F19F24.9  
T01617 Length: 392 February 11, 2000 15:52 Type: P Check: 1414 ..  
1 MKYIRNSLK RLFSFKRSF DSDSENSSP LAASSTKVE GFQETDOFOR  
51 PKMKCFQOE IYDATINGESS ENLVGRGGA EYKGIIGNN GEELAVKRIT  
101 RGRGRDERRE KEFLMEIGTI GHVSHPNVLS LLGCCINDGL YLVFISSRG  
151 SIASLHDLN QAPLEWETRY KIAIGTAKGL HYLHRCQQR IHHDIKSSN  
201 VLLNDFEPQ ISDPGLAKWL PSQMSHNSIA PIEGTFGHLA PEYTHGIYD  
251 EKTIVFARCV FLELISGKK PYDASHQSLH SWAKLIIDKG EIEKLVDPRI  
301 GEEDLQQLH RIAFASICI RSSSLCRPSM IEVLVLOGE DIEKEMWKE  
351 EEEVKEEFW GYEDLEDCEC DSSISLSPD SISNRSSSHR SR  
!!AA\_SEQUENCE 1.0  
P1:T06062 - hypothetical protein F19H22.120 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse ear cross)  
C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 30-Apr-1999  
C:Accession: T06062  
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schueller, C.  
submitted to the Protein Sequence Database, March 1999  
A:Reference number: 215184  
A:Accession: T06062  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-169 <BEV>  
A:Cross-references: EMBL:AL035679  
A:Experimental source: cultivar Columbia; BAC clone F19H22  
C:Genetics:  
A:Map position: 4  
A:Note: F19H22.120  
T06062 Length: 169 February 11, 2000 15:52 Type: P Check: 59 ..  
1 MYSKPIGCF AFRRHQHLL LRHGIHSOK EBNFLQVDE GGMIVPLIT  
51 IYVTKIIMM ACCCECCQ MEDDAYSEWL RANGVNAS QNHNGSDAM  
101 GYFLGEVWFP YQADSDFELS LSVGDYVIR EVVSVWMEG ECKNGAMFT  
151 YYIERDRV FAKYIEVF  
!!AA\_SEQUENCE 1.0  
P1:T04272 - hypothetical protein F20B18.280 - Arabidopsis thaliana (fragment)  
C:Species: Arabidopsis thaliana (mouse ear cross)  
C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 30-Apr-1999  
C:Accession: T04272  
R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hohnselt, J.; Mewes, H.W.; Mayer, K.F.X.; Schueller, C.  
submitted to the Protein Sequence Database, March 1999  
A:Reference number: 215263  
A:Accession: T04272  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-349 <BEV>  
A:Cross-references: EMBL:AL049483  
A:Experimental source: cultivar Columbia; BAC clone F20B18  
C:Genetics:  
A:Map position: 4  
A:Note: F20B18.280  
T04272 Length: 349 February 11, 2000 15:52 Type: P Check: 5780 ..  
1 KQILIGANER ENFREKDLV GNRVQGAFO GLVELSHDG RRDVLVANTL  
51 GQPESTRRL RSYGRFAHH DLLKGLISQT ILPTQKSD NOTEEKKSDS  
101 EEEREVSSDA AEKESNSLPS ILRLSRQO PVSEKHDDIV DESDSASAG  
151 VLEEDGTCT TTPYGRKRC TEHKGRLSR VSPDIHICE VPTVRECEET

Mon Feb 14 08:07:32 2000

pir.cat

Page 160

```
201 ENIGVILPD MIRCRSPVS RRRCEBHG MRVNAEFFLL NPTERKAVN
251 EDKRPETST GANQEGSGL CEATTKNGLP CTRSAPEGK RCMQHKDRL
301 NHGSENVQS ATRSOVIGCF KLYNGSVCEK SPVGRKRCE EHKMRITS

!!A_SEQUENCE 1.0
P1:T05639 - hypothetical protein F20D10.240 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
C:Accession: T05639
R:Bevan, M.; Medler, H.; Kutzner, M.; Wambutt, R.; Bancroft, I.; Mewes, H.W.;
Mayer, K.F.X.; Scheller, C.
Submitted to the Protein Sequence Database, February 1999
A:Reference number: Z15420
A:Accession: T05639
A:Molecule type: DNA
A:Residues: 1-445 <BEV>
A:Cross-references: EMBL:AL035538
A:Experimental source: cultivar Columbia; BAC clone F20D10
C:Genetics:
A:Map position: 4
A:Introns: 88/3; 148/2; 234/3; 283/1; 399/3
A:Note: F20D10.240

T05639 Length: 445 February 11, 2000 15:52 Type: P Check: 5981

1 MFTAAASSV GRMRTAFSL RDEISTTPP PVPLLEDL FQSHSLISA
51 VSHLPHLHET SDCLFLDLV SKADGPDWIP VSRHTCOLIH DVCARLLFOL
101 NSSMPLLIH SFASVLEFL RQPMSPSPS AAFVSRIEPLV IQCFETLRL
151 APMHEPHL VKFLVRVPL LHQDLVLSYG FSNODESPFL IVEKRLPOON
201 KLMSMALAF DMFGRAPSL ESLEPTDVSQ CTEVLARKVA DVLASKGLV
251 EDREMRWMP LVMLRLOFTF FFLGSIRLVA LLASLMFPC FGLTGPPLS
301 VSDVYHNDK LNVKISPLIS GVSNAKNTP YRPPHRRKD DLNTRPVSS
351 SWRLSAHDS GSSDVISDS DFSDSDGSVP DSYFAQSKV RIAIVCIQD
401 LQADSKSFT TQWTLFPTS DVLKPRKIN LQHLSPYS IFLSP

!!A_SEQUENCE 1.0
P1:T01144 - hypothetical protein F26B6.20 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 30-Apr-1999
C:Accession: T01144
R:Roundley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon,
Venter, J.C.
Submitted to the EMBL Data Library, June 1998
A:Description: Arabidopsis thaliana chromosome II BAC F26B6 genomic sequence.
A:Reference number: Z14198
A:Accession: T01144
A:Status: translated from GB/EMBL/DBJ
A:Residues: 1-272 <ROU>
A:Molecule type: DNA
A:Cross-references: EMBL:AC003040; NID:93242700; PID:93242730
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 2
A:Introns: 131/3; 177/3
A:Note: F26B6.20

T01144 Length: 272 February 11, 2000 15:52 Type: P Check: 6323

1 MLVRKAMEN KNORFVLIH GVCHGAMTWK KVKTOLEVAG HCVTVNDLAA
51 SGIMTKVEE IOTLNDYCKP LLEFSLIGS DDGKVIYVAH SMGGSALAA

101 ADFACKIAA IYFLAEMPD TINPAYVE KILRSIPOE WLDTCVNG
151 KPDPFLQYTL LGFKEMAKM YONSPVODLE VYKTLVRENP LVTNLAGTR
201 SFSEGGVSV TRIIVYCRD LVEVEDYORW MISNPPREV MEIKADHP
251 MFSKPOEVA LLEIANKYC KN

!!A_SEQUENCE 1.0
P1:T01145 - hypothetical protein F26B6.21 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 30-Apr-1999
C:Accession: T01145
R:Roundley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon,
Venter, J.C.
Submitted to the EMBL Data Library, June 1998
A:Description: Arabidopsis thaliana chromosome II BAC F26B6 genomic sequence.
A:Reference number: Z14198
A:Accession: T01145
A:Status: translated from GB/EMBL/DBJ
A:Residues: 1-260 <ROU>
A:Molecule type: DNA
A:Cross-references: EMBL:AC003040; NID:93242700; PID:93242731
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 2
A:Introns: 124/3; 167/3
A:Note: F26B6.21

T01145 Length: 260 February 11, 2000 15:52 Type: P Check: 9711

1 MKNOKKFTV LVHIGCHAW CWYKVAOLE AAGHSTAVD LAASGVNNTS
51 LDEIOTLDY CKPLLEFLS LGSDDKVIL VASMGISLA SLADIPFSK
101 VAAIVFAAF MPDISNPAY VFQKLVKDYV QEVMQDTVG KPDPLEFAL
151 FGPEFMAYTL YNLSPLDPE LAKMSVRVSP FMTNLAGTI SFSDRGVSV
201 TRIIVYGED VAVPYDORG MINDFPVKEV LEIKADHP MFSRPELCA
251 LLEIADKXA

!!A_SEQUENCE 1.0
P1:T08549 - hypothetical protein F27B13.120 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C:Accession: T08549
R:Bevan, M.; Zimmermann, W.; Gruenelsen, A.; Wambutt, R.; Bancroft, I.; Mewes,
H.W.; Mayer, K.F.X.; Scheller, C.
Submitted to the Protein Sequence Database, May 1999
A:Reference number: Z16442
A:Accession: T08549
A:Status: preliminary
A:Residues: 1-404 <BEV>
A:Molecule type: DNA
A:Cross-references: EMBL:AL050352
A:Experimental source: cultivar Columbia; BAC clone F27B13
C:Genetics:
A:Map position: 4
A:Introns: 89/3
A:Note: F27B13.120

T08549 Length: 404 February 11, 2000 15:52 Type: P Check: 4270

1 MICEPAYOYQ QHGQKDHMM TMAMDLS TS PPSSPLSPSL SPKSSYNNN
51 EERLEVNL SGMALQLPN PSLNLANICK LIDSNNHIK IPESITALL
101 NLIADISHN QIKALPSIG CISKILINV SGNFVSLQ TIQNCRFTS
```

```

151 TSGGNQSM MIMINSVF FCGFGRSLE ELNANENELI RLPDNGIEL
201 TNLKLCVNS NKILSPATI TGLTSRVID ARUCLMILP EDLENINLE
251 LNVSONFOY LSLPSSIGL LNLLELDS YNKITVLES ICGMRRLRL
301 SAEGNPLVSP PIEVEQSLH AVREYSQMA NGKLVNNAK KTWGRKRLV
351 KYGTENGSR VWTREEREL IMPEYRPIDI LSTKFPVTC SPRRLSPRT
401 YFSR

!!A_SEQUENCE 1.0
P1:T05096 - hypothetical protein F28M20.70 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C:Accession: T05096
R:Bevan, M.; Rieger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Hohnescl,
J.; Mewes, H.W.; Mayer, K.F.X.; Schueller, C.
submitted to the Protein Sequence Database, November 1998
A:Reference number: Z15398
A:Accession: T05096
A:Molecule type: DNA
A:Residues: 1-171 <BEV>
A:Cross-references: EMBL:AL031004
A:Experimental source: cultivar Columbia; BAC clone F28M20
C:Genetics:
A:Map position: 4
A:Introns: 30/1
A>Note: F28M20.70

T05096 Length: 171 February 11, 2000 15:52 Type: P Check: 9032 ..

1 MDKLFALMY LLSLETINOS EVEAEALP SADSAKRSNI YDMAEKLTYQ

51 SISAVTPSVK NLSSDOQLP VARYEALTD GRPNPNDST IILDAPASKS
101 GSIDNSYVK GPEBAIVFM GGNITIESS LQELSQOEM VNNITGATE
151 ILTGELVEQ LGELGQRMGL R

!!A_SEQUENCE 1.0
P1:T02589 - hypothetical protein T16B24.23 - Arabidopsis thaliana (fragment)
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 23-Apr-1999
C:Accession: T02589
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon,
R.C.; Sykes, S.M.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.;
Venter, J.C.
submitted to the EMBL Data Library, August 1998
A:Description: Arabidopsis thaliana chromosome II BAC T16B24 genomic sequence.
A:Reference number: Z14679
A:Accession: T02589
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-381 <ROU>
A:Cross-references: EMBL:AC004697; NID:93402671; PID:93402690
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 2
A:Introns: 60/2, 169/1, 240/3, 248/3
A>Note: T16B24.23

T02589 Length: 381 February 11, 2000 15:52 Type: P Check: 3673 ..

1 MNHVPSQSF YIESEDEDR KDYYEDDGS HSDSSDDYD ENQAHKPS
51 YTTAMPQSYR QSIDLYSSVP SPGIGFLGN SMTRFSSFL SSGILRRATP
101 ESLPTVKPL LEOADEQAL PKHRLSSGL LSPIPRRRS MRKCKSSMV
151 SHEIPMSRNS SYGQAVLNGI NVLCGVGIS TPTAAGGM LGIMLFYTG

!!A_SEQUENCE 1.0
P1:T02510 - hypothetical protein T19C21.16 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 14-May-1999
C:Accession: T02510
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,
S.M.; Kaul, S.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.;
Venter, J.C.
submitted to the EMBL Data Library, August 1998
A:Description: Arabidopsis thaliana chromosome II BAC T19C21 genomic sequence.
A:Reference number: Z14676
A:Accession: T02510
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-96 <ROU>
A:Cross-references: EMBL:AC004683; NID:93395421; PID:93395437
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 2
A:Introns: 17/3
A>Note: T19C21.16

T02510 Length: 96 February 11, 2000 15:52 Type: P Check: 8930 ..

1 MSRYLRSHA WCLRQNVSN KSFLEKSFV GLGYTPPIRY FIDIGPLPMW

51 TRKEVNEILL GRTRVSPIL KLILFLVHR ICSDPRFOVK SSSVNA

!!A_SEQUENCE 1.0
P1:T0212 - hypothetical protein T20B5.17 - Arabidopsis thaliana
N:Alternate names: hypothetical protein T3K9.26
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 14-May-1999
C:Accession: T00761; T02122
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,
S.M.; Kaul, S.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.;
Venter, J.C.
submitted to the EMBL Data Library, November 1997
A:Description: Arabidopsis thaliana chromosome II BAC T20B5 genomic sequence.
A:Reference number: Z14159
A:Accession: T00761
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-248 <ROU>
A:Cross-references: EMBL:AC002409; NID:92623294; PID:92623311
A:Experimental source: cultivar Columbia
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon,
R.C.; Sykes, S.M.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.;
Venter, J.C.
submitted to the EMBL Data Library, February 1999
A:Description: Arabidopsis thaliana chromosome II BAC T3K9 genomic sequence.
A:Reference number: Z14570
A:Accession: T02122
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-248 <ROU>
A:Cross-references: EMBL:AC004261; NID:93402695; PID:93402720
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 2
A>Note: T20B5.17; T3K9.26

T02122 Length: 248 February 11, 2000 15:52 Type: P Check: 8329 ..

1 MREDNPMFL RMEELPSPB ELIPISQTLI TPTLALAFQI GSPNHHGSK

```

```

51 RTAIYHQKL QSSTPTPTPT PTPPMAMNS DFGGSDTDL GSGSIGEPA
   P1:T02348 - hypothetical protein T26B15.5 - Arabidopsis thaliana
   C:Species: Arabidopsis thaliana (mouse-ear cress)
   C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 23-Apr-1999
   C:Accession: T02348
   R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.
   Submitted to the EMBL Data Library, July 1998
   A:Description: Arabidopsis thaliana chromosome II BAC T26B15 genomic sequence.
   A:Reference number: Z14678
   A:Accession: T02348
   A:Status: translated from GB/EMBL/DBJ
   A:Molecule type: DNA
   A:Residues: 1-217 <ROU>
   A:Cross-references: EMBL:AC004681; NID:g3298532; PID:g3298537
   A:Experimental source: cultivar Columbia
   C:Genetics:
   A:Map position: 2
   A:Introns: 25/1
   A>Note: T26B15.5

T0548 Length: 217 February 11, 2000 15:52 Type: P Check: 5743 ..

1 MKRGKHLCP NGFTGYSSIH HHYEHNVDF FEERLIVYT HTPSVIRMT
   P1:T05859 - hypothetical protein T29A15.30 - Arabidopsis thaliana
   C:Species: Arabidopsis thaliana (mouse-ear cress)
   C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 13-Aug-1999
   C:Accession: T05859
   R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Hohelsel, J.; Mewes, H.W.; Mayer, K.F.X.; Schueller, C.
   Submitted to the Protein Sequence Database, March 1999
   A:Reference number: Z15455
   A:Accession: T05859
   A:Molecule type: DNA
   A:Residues: 1-208 <BEV>
   A:Cross-references: EMBL:AL035602
   A:Experimental source: cultivar Columbia; BAC clone T29A15
   C:Genetics:
   A:Map position: 4
   A:Introns: 151/2
   A>Note: T29A15.30

T05859 Length: 208 February 11, 2000 15:52 Type: P Check: 7427 ..

1 MAFSPPLSL SVDPAPFESW LRDSGYIELL DHRTSAAAA ASSASVSSS
   51 AAATSAASD VVSTITGGEF ASLISRLVTV SGLTINFS KLSADDESGD
   101 TTPWTGFIG NCDYSFPSS SQOARRVHE NIKRPARYA TLTIVFACA
   151 LYQMPALVG LIGSLAMEL FKYSCKWKF DRHPSMRKLS IGIGOCGEFF
   201 DHEVPVFC

11AA_SEQUENCE 1.0
P1:T01365 - hypothetical protein T29F13.9 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 30-Apr-1999
C:Accession: T01365
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.
Submitted to the EMBL Data Library, May 1998
A:Description: Arabidopsis thaliana chromosome II BAC T29F13 genomic sequence.
A:Reference number: Z14177
A:Accession: T01365
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-175 <ROU>
A:Cross-references: EMBL:AC003096; NID:g3132469; PID:g3132475
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 2
A:Introns: 73/1
A>Note: T29F13.9

T01365 Length: 175 February 11, 2000 15:52 Type: P Check: 1733 ..

1 MGLVTKALKY LLLSLSLF IYTNVSSAS PMTPSSPAK MSRLVAVEG
   51 MYCKSCSKYS GVDTLLEASP LOGATVKLAC NNTKRGVME TKTDKNGYFF
   101 MIAPKKITY AFHTCRAMPT NPGPTTATMT CTYPSKLNNG ITGAMLRPSK
   151 TINIGHDYV LFSVGPFAE PACAL

11AA_SEQUENCE 1.0
P1:T08903 - hypothetical protein T32A16.10 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C:Accession: T08903
R:Bevan, M.; Zimmermann, W.; Gruenelsen, A.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Lemcke, K.; Schueller, C.
Submitted to the Protein Sequence Database, May 1999
A:Reference number: Z16518
A:Accession: T08903
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-364 <BEV>
A:Cross-references: EMBL:AL078468
A:Experimental source: cultivar Columbia; BAC clone T32A16
C:Genetics:
A:Map position: 4
A:Introns: 153/1; 325/1
A>Note: T32A16.10

T08903 Length: 364 February 11, 2000 15:52 Type: P Check: 2816 ..

1 MNTCTIVSEP KTHSSLASLK KLVLSGANFS AETESISFTN KSCITYLDVS
   51 KTSLSKNSFL ETWFNLEHLD LSTAFGDDS VGFVACVGEN LKNLVSDTQ
   101 ITPGVGNLA GHVPOLETLS MSQTFVDDL S ILLISTWPC IKALDLGNNS
   151 TLGFYILISP QEEKEKSLLA LQSLISLETI SLEHPTLGDK ALSGLSSLTG
   201 LTHSLTSTS LNDSTLHLS SLPLVSLGV RQCVLTISNCL EKFRPNRLR
   251 TLDIQGCWML TKDDIAGLCK RYPIKVRHE HADSSLDON QLDRSSTPQ
   301 SFGVYARRNN QRPESVAVP RSFLDORVKY NREELVALON SPLSLPWE
   351 ELVSVPELLA DSVI

11AA_SEQUENCE 1.0
P1:T06094 - hypothetical protein T5J17.50 - Arabidopsis thaliana

```

C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 30-Apr-1999  
 A:Accession: T06094  
 R:Bevan, M.; Murphy, G.; Riddle, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.;  
 Mayer, K.F.X.; Schueller, C.  
 submitted to the Protein Sequence Database, March 1999  
 A:Reference number: 215184  
 A:Accession: T06094  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-178 <BEV>  
 A:Cross-references: EMBL:AL035708  
 A:Experimental source: cultivar Columbia; BAC clone T5J17  
 C:Genetics:  
 A:Map position: 4  
 A:Note: T5J17.50

T06094 Length: 178 February 11, 2000 15:52 Type: P Check: 7672

1 MGSRLGGRV HFANLPKIL MPKLTNIHE FALKTIPSAS KIEIKRVLES  
 51 LYGFDEKYN TLNDGKKKK RGGLIAKAD YKRAYTLRS PLISRDLPF  
 101 VKTIEDRKS KYKGSFVEE EDDKSHMLD RREKREIGGY GKSGRRGG  
 151 RANSPTRGGA AAGTAKFPW SNMRFVCK

11AA\_SEQUENCE 1.0  
 P1:T00988 - hypothetical protein T9J22.22 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 14-May-1999  
 R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,  
 S.M.; Kaul, S.; Mason, T.M.; Kellavage, A.R.; Adams, M.D.; Somerville, C.R.;  
 Venter, J.C.  
 submitted to the EMBL Data Library, April 1998  
 A:Description: Arabidopsis thaliana chromosome II BAC T9J22 genomic sequence.  
 A:Reference number: Z14153  
 A:Accession: T00988  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-241 <ROU>  
 A:Cross-references: EMBL:AC002505; NID:92739359; PID:92739380  
 A:Experimental source: cultivar Columbia  
 C:Genetics:  
 A:Map position: 2  
 A:Introns: 159/1  
 A:Note: T9J22.22

T00988 Length: 241 February 11, 2000 15:52 Type: P Check: 916

1 MASLAPPL LSTPRKLTLS HLHTSISPF QISTORRPOK HLINICRSTP  
 51 TPQQAASOR KRTYRKQYP GENIGITEEM RFVAMRLNV NGKIDLSGD  
 101 KITEKEEEE EEDDDDDDE VKDETWRPSK EGFLLKLYDS KLVEFTIERI  
 151 VDESENVSTA YFRRTGLEERC ESEKDLWL REODVLEP SNVGSYARY  
 201 LEOAGESAP LFLSHFYSIV FSHIAGGYL VQVSEFFGL S

11AA\_SEQUENCE 1.0  
 P1:T08191 - probable translation elongation factor ts - Arabidopsis thaliana  
 N:Alternate names: protein T22B4.100  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 11-Jun-1999  
 A:Accession: T08191  
 R:Bevan, M.; Pohl, T.; Weizenecker, T.; Bancroft, I.; Mewes, H.W.; Mayer,  
 K.F.X.; Schueller, C.  
 submitted to the Protein Sequence Database, May 1999  
 A:Reference number: Z16098  
 A:Accession: T08191  
 A:Status: preliminary

A:Molecule type: DNA  
 A:Residues: 1-415 <BEV>  
 A:Cross-references: EMBL:AL049876  
 A:Experimental source: cultivar Columbia; BAC clone T22B4  
 C:Genetics:  
 A:Map position: 4  
 A:Introns: 104/1; 186/3; 216/3; 275/1; 331/3; 373/3; 403/2  
 A:Note: T22B4.100

T08191 Length: 415 February 11, 2000 15:52 Type: P Check: 2415

1 MAPARVRRP IGVYIVSVS RRSNGEYST VASKETLSQ YKSVPSGCT  
 51 SLVRFGNFI RSESEAPPA VSDQSLIKQ LRERTSAPIK DYKASLVCN  
 101 WDLVFFGFD TEAAOKDLK RGVLAASKS SRTAEGLA VQNGKAVV  
 151 IELNCTDFV ARNEIFQYL SPALSHWLL LNDGLFLAM AKHALVESS  
 201 SQGVSVFPE GPDLFEFEKL NUDPKVNGE TTVSNAVTEV AAIMGNVAF  
 251 RRGFLMSKS AGVLAVALT SPQAGIVSL EVEGNTOLE AIORVSELA  
 301 MHVAAKPLF LSKDLVSSA MANREILKS QASTGNOM AIEKIYEGRL  
 351 RKFEEVALM EQKEFYNDAL NIKTLVDNIS KEVSPKAVT DELRVEGEC  
 401 IERLEADEP VAQTA

11AA\_SEQUENCE 1.0  
 P1:T09662 - multicatalytic endopeptidase complex chain Y13 homolog - alfalfa  
 (frugent)  
 N:Alternate names: proteasome chain Y13 homolog  
 C:Species: Medicago sativa (alfalfa)  
 C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 16-Jul-1999  
 R:Frings, G.  
 submitted to the EMBL Data Library, April 1996  
 A:Description: Isolation of an alfalfa DnaU-like gene and a Y13 proteasome  
 subunit homologous gene showing differential expression in embryogenic and  
 non-embryogenic callus  
 A:Reference number: Z16763  
 A:Accession: T09662  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-101 <FRU>  
 A:Cross-references: EMBL:Z71998; NID:e992787; PID:e239056  
 A:Experimental source: variety Kangelande  
 C:Keywords: hydrolase; proteinase

T09662 Length: 101 February 11, 2000 15:52 Type: P Check: 3889

1 CDPSGNYGWM KAGAIANNQ AAOSILKODY KDDITREAV NLALVLSKI  
 51 MDRTSLTSDK LELAELIAP SKRVKYVCS PENITKLIVK SGVTPAIEET  
 101 R

11AA\_SEQUENCE 1.0  
 P1:S23774 - triose phosphate/3-phosphoglycerate/phosphate translocator  
 precursor - garden pea  
 C:Species: Pisum sativum (garden pea)  
 C:Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 09-Sep-1997  
 R:Willege, D.L.; Fischer, K.; Wachter, E.; Link, T.A.; Fluegge, U.I.  
 Plantae 183: 451-461, 1991  
 A:Title: Molecular cloning and structural analysis of the phosphate  
 translocator from pea chloroplasts and its comparison to the spinach phosphate  
 translocator.  
 A:Reference number: S23774  
 A:Accession: S23774  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA

A/Residues: 1-402 <WIL>  
 A/Accession: S23816  
 A/Molecule type: protein  
 A/Residues: 73-95;97-102;104;106;108-115 <WIL1>  
 A/Note: 227-leu was also found  
 R.Schnell, D.J. Blodet, G.; Pain, D.  
 J. Cell Biol. 111, 1825-1838, 1990  
 A/Title: The chloroplast import receptor is an integral membrane protein of chloroplast envelope contact sites.  
 A/Reference number: A37791; M01D:91035597  
 A/Accession: A37791  
 A/Molecule type: mRNA  
 A/Residues: 1-402 <SCH>  
 A/Cross-references: EMBL:X54639; NID:920690; PID:920691  
 A/Accession: S23775  
 A/Molecule type: protein  
 A/Residues: 'X', 74-98;249-260;269-289;324-350 <SCH1>  
 A/Note: this protein was identified as chloroplast import receptor  
 C/Keywords: chloroplast; transmembrane protein  
 E/1-72/Domain: transit peptide (chloroplast) #status experimental <TNP>  
 E/73-402/Product: triose phosphate/3-phosphoglycerate/phosphate translocator #status experimental <MAT>  
 E/97-117/Domain: transmembrane #status predicted <TM1>  
 E/129-151/Domain: transmembrane #status predicted <TM2>  
 E/168-190/Domain: transmembrane #status predicted <TM3>  
 E/214-234/Domain: transmembrane #status predicted <TM4>  
 E/273-292/Domain: transmembrane #status predicted <TM5>  
 E/342-364/Domain: transmembrane #status predicted <TM6>  
 F/367-388/Domain: transmembrane #status predicted <TM7>  
 S23774 Length: 402 February 11, 2000 15:52 Type: P Check: 5889 ..

1 MESRVLSRAT TLSSLPITLKK LHRPLPLANAS LPSVSGFSV SDGGLWGR  
 51 OLRLPELCSPV LKKGASLLRP CPATAGGNDG AGEKVAEFG FFSRYPALTT  
 101 GFEFTWYFL NVIPNILLKK IYVPPYFV VSVTHLVGV VGLVSWTVG  
 151 LPRRAIDGN LKLLIPAV CHALGHVTSN VSFAVAVSF THTVKALEPF  
 201 FNAAAGQFL GOSIPITLML SLAPVYIGVS MASLTELSPN WGFISAMIS  
 251 NISFTYRSY SKRAMTDMDS TNYIVYSII ALIYCIPPAL IIEGPTLLKT  
 301 GRNDALAKVG LVKFSDFLW VGMFYHLNQ VAIWTLERVA PLTHAVGNVL  
 351 KRVFVIGFSI IIFGNKISTO TGIGTGIAIA GVALYSIKA QIEEKRQAK  
 401 AA

!!AA\_SEQUENCE 1.0  
 P1:S22500 - embryonic protein ecp31 - carrot  
 C/Species: Daucus carota (carrot)  
 C/Date: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 09-Sep-1997  
 A/Accession: S22500  
 R.Kiyosue, T.; Yamaguchi-Shinozaki, K.; Shinozaki, K.; Higashi, K.; Satoh, S.; Kamada, H.; Harada, H.  
 Plant Mol. Biol. 19, 239-249, 1992  
 A/Title: Isolation and characterization of a cDNA that encodes ECP31, an embryogenic-cell protein from carrot  
 A/Reference number: S22500; M01D:92322951  
 A/Accession: S22500  
 A/Molecule type: mRNA  
 A/Residues: 1-256 <KIT>  
 A/Cross-references: EMBL:X60593; NID:918336; PID:918337  
 C/Genetics:  
 A/Genes: ecp31  
 S22500 Length: 256 February 11, 2000 15:52 Type: P Check: 9347 ..

1 MSQQQRRRQ QEQPIKGDV EDVSGQLSSQ FVAPNDASAM QAAENNVLAK  
 51 TQKGASVAY OSAAANLQR GVVPHEGTP IASEGVATIS EAEIATRRI

101 TEAVGGQVVG QYLEPGKFKK PSPAGVLGSD SITIGEALET TALTAGDKPV  
 151 DOSDAAAIQA AEVRASGYAY PGVAAAAS AADYNARTMN VASTKXGCV  
 201 IADRSIRLAE DKAVTREDAE GVGAERYNN PEMMTYPGV ASSMAAARL  
 251 NODPTF

!!AA\_SEQUENCE 1.0  
 P1:T16984 - transcription factor homolog Brf3 - curled-leaved tobacco  
 C/Species: Nicotiana glauca (curled-leaved tobacco)  
 C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
 A/Accession: T16984  
 R.Borisjuk, N.V.  
 submitted to the EMBL Data Library, October 1996  
 A/Reference number: Z18621  
 A/Accession: T16984  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-165 <BOR>  
 A/Cross-references: EMBL:Y09106; NID:91008749; PID:9280517  
 A/Experimental source: somatic embryo  
 C/Keywords: transcription factor  
 T16984 Length: 165 February 11, 2000 15:52 Type: P Check: 4925 ..

1 MNVEKLRKKA GSVRTGKGT MRRKKRAVHK TTTTDRRLQ STLKRIQVNA  
 51 IPAIEEVNIF KEDVYIOFIN PKVQASIAN TWVYSGSPQT KTLQGYSSNN  
 101 YSPVGDNIE SLREASRAVP ESRASANGA PEGALALQED DDEVEPELVA  
 151 GQTFEAGNSE ERHTS

!!AA\_SEQUENCE 1.0  
 P1:S47086 - pir7a protein - rice  
 C/Species: Oryza sativa (rice)  
 C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 09-Sep-1997  
 A/Accession: S47086  
 R.Reilmann, C.; Mauch, F.; Dudler, R.  
 submitted to the EMBL Data Library, June 1994  
 A/Description: Host-response of rice to resistance-inducing agents: characterisation of a gene specifically regulated by infiltration of pseudomonas syringae pv. syringae.  
 A/Reference number: S47085  
 A/Accession: S47086  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-263 <REI>  
 A/Cross-references: EMBL:Z34271; NID:9498743; PID:9498744  
 C/Genetics:  
 A/Introns: 122/3; 168/3  
 S47086 Length: 263 February 11, 2000 15:52 Type: P Check: 1305 ..

1 MEDGKRFVE VAGLGYGMC WYRVVAALRA AGHRAMALDM AAGAPPARA  
 51 DEVGSLFYS RPLDANVAA APGERLVVG HSLGLSLAL AMRFEPDKVA  
 101 AAYFLACNP AAGKHGITT EEFMRIRKP FFMDSKTIIVL NTNQEPRTAV  
 151 LLPKLIARK LYNRPPEDL TLATMLVRPG TNYIDDDIMK DELLTEGNY  
 201 GSVKRVFLVA MDDASDDEM ORWTIDLSPG VEVELAGAD HNAKSKFRE  
 251 LQDLIRIAA KYD

!!AA\_SEQUENCE 1.0  
 P1:S14884 - hypothetical protein 7 - yeast (Hansenula polymorpha)  
 C/Species: Hansenula polymorpha  
 C/Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Sep-1997  
 A/Accession: S14884



R:Krutillina, A.I.; Seragina, S.A.; Tikhomirova, L.P.; Kryukov, V.M.  
 submitted to the EMBL Data Library April 1991  
 A:Description: Nucleotide sequence of Hansenula polymorpha DNA region  
 Complementing DNA.  
 A:Accession: S14854  
 A:Reference number: S14854  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-146 <KRU>  
 A:Cross-references: EMBL:X58862; NID:92762; PID:92769

S14884 Length: 146 February 11, 2000 15:52 Type: P Check: 7397

1 MGRTSGGIYS IFISALAKSL KDREIQGGY EYTPPELLAAS LKDALDSLYR  
 51 YTRARAGDRT LIDALAPFE GPAAKSGDLN RANKACHEGA ESTRLAKAF

101 GRASYSESE FKPEAEAGGL PDPGALGLAA LVDFAPAEAS KIGSNL

!!NA\_SEQUENCE 1.0  
 P1:S43448 - FUN54 protein precursor - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: probable membrane protein YAL007c  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 13-Jan-1995 #sequence\_revision 10-Feb-1995 #text\_change 17-Mar-1999  
 C:Accession: S43448; S40896  
 R:Clark, M.W.; Kemp, T.; Storms, R.K.; Zhong, W.; Fortin, N.; Zeng, B.;  
 Delaney, S.; O'Neill, B.F.F.; Barton, A.B.; Kaback, D.B.; Bussey, H.  
 Yeast 10, 535-541, 1994  
 A:Title: sequencing of chromosome I of Saccharomyces cerevisiae: analysis of  
 the 42 kbp SPOT-CENT-ODC15 region.  
 A:Reference number: S43448; MUID:95028152  
 A:Accession: S43448  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-213 <CL>  
 A:Cross-references: EMBL:L22015; NID:91339990; PID:9349746; MIPS:YAL007c  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library,  
 November 1993

C:Genetics:  
 A:Gene: FUN54  
 A:Cross-references: MIPS:YAL007c

A:Map position: 1L  
 C:Keywords: glycoprotein; transmembrane protein  
 F:1-25/Domain: signal sequence #status predicted <SIG>  
 F:26-215/Product: FUN54 protein #status predicted <AMT>  
 F:183-203/Domain: transmembrane #status predicted <TM>  
 F:170-175/Binding site: carbohydrate (asn) (covalent) #status predicted

S43448 Length: 215 February 11, 2000 15:52 Type: P Check: 5386

1 MIKSTIALPS FFIYLILALV NSVAASSYA PVALISPARS KECLYIDMYT  
 51 EDDSLAVGYQ VLTGNFEID FDIAPDGSV ITSEKOKKIS DILKSFQVG

101 KTFCEFSNNY GIALKVEIT LEKEKTLTDE HEADVNMNDI INNAVEEID  
 151 RNLNITITL NYLRAREMRN MSTVNSTESR LTMLSILITL ITAVISIAY

201 LLIOELFTGR OKNYV

!!NA\_SEQUENCE 1.0  
 P1:S19425 - hypothetical protein YCR015c - yeast (Saccharomyces cerevisiae)  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 14-Nov-1997  
 C:Accession: S19425  
 R:Batat, D.; Jacq, C.; Perez, J.; Shu, Y.  
 Submitted to the Protein Sequence Database, March 1992  
 A:Reference number: S19424  
 A:Accession: S19425  
 A:Molecule type: DNA  
 A:Residues: 1-317 <HAT>  
 A:Cross-references: EMBL:X59720; NID:91907115; PID:e264479; PID:91907158;  
 MIPS:YCR015c

C:Genetics:  
 A:Map position: 3R  
 S19425 Length: 317 February 11, 2000 15:52 Type: P Check: 1969

1 MKTIIDDFD EITRVDTIC TIAPLYLLN PRLKPEWHF TKTYMDGYHK  
 51 KYKGTSLP LSSGVPTII SOSNFKLFA DELKYNHNR VELSVMET  
 101 TKQITFSIS LQMKTFARD QNHDDCLRD GERTCSSVY KNEBDFYVL  
 151 SINMSKEPIH EYIGDRRLKN SHIFCNLDR VSDKCSOSYN GEFCRLITG  
 201 SDRKALIGET LDKIDSGCKR EGNCSYWI GDEETDLISI LHPSTNGVLL  
 251 INPOENPSKE IKITEKILGI PKDKISSFEA DNGPAVLQFC EKEGKGAVL  
 301 VKSWDSLKDL IMQVTKM

!!NA\_SEQUENCE 1.0  
 P1:S70131 - hypothetical protein YDR273w - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: hypothetical protein D9954.4  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 24-Aug-1996 #sequence\_revision 06-Sep-1996 #text\_change 12-Dec-1997  
 C:Accession: S70131  
 R:Le, T.  
 Submitted to the EMBL Data Library, May 1996  
 A:Description: The sequence of S. cerevisiae cosmid 9954.  
 A:Reference number: S70124  
 A:Accession: S70131  
 A:Molecule type: DNA  
 A:Residues: 1-355 <LET>  
 A:Cross-references: EMBL:U51030; NID:91332633; PID:91230644; MIPS:YDR273w  
 C:Genetics:  
 A:Map position: 4R  
 A:Note: YDR273w

S70131 Length: 365 February 11, 2000 15:52 Type: P Check: 1796

1 MGKKNRKGKE NNAKTSFLK VENIKNTNGO IELPSQDYTN VERKESSPKT  
 51 DFPLITKEHV NTKTDSNILD YPTIGDLVSS VEKLVKLKL KIAPEYVDY  
 101 LKAILLASQ GYLEPAFNSL LYSSPEENT DELAPMKRPS VEDYSKINVS  
 151 EILQREIFDD IDEPFGGQI NGSMVISKIE SETSLIAEHI GNISTPGSNR  
 201 EVAESTRNVA VABGHNTILS NEDSILKGE KGKEEKEKG EEKGVNSLKG  
 251 AAVKYVAKSL KNNRIPYTK RNEPSNNLFD VLNCDSESE EEOOVETNTS  
 301 NOERNKOGN TEVPEAKORDS ADRLPAKDDG GYKSAFGTDS CGLFAADAKD  
 351 ERKQVHPSRQ ELSEF

!!NA\_SEQUENCE 1.0  
 P1:S69699 - hypothetical protein YDR415c - yeast (Saccharomyces cerevisiae)  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 22-Aug-1996 #sequence\_revision 06-Sep-1996 #text\_change 21-Nov-1997  
 C:Accession: S69699  
 R:Dietrich, F.S.  
 Submitted to the EMBL Data Library, August 1995  
 A:Description: The sequence of S. cerevisiae lambda 3641 and cosmids 9461,  
 9831, and 9410.  
 A:Reference number: S69553  
 A:Accession: S69699  
 A:Molecule type: DNA  
 A:Residues: 1-374 <DIE>  
 A:Cross-references: EMBL:U03007; NID:927685; PID:9927713; MIPS:YDR415c  
 C:Genetics:  
 A:Map position: 4R

S69699 Length: 374 February 11, 2000 15:52 Type: P Check: 3355

Mon Feb 14 08:07:32 2000

plr.cat

Page 166

```
1 MROSLFVLF NVAIANSYP YEPLRLVQV ENEMVEPES EKLNRGRG
51 KEFDVTKHTS FLPEFNKEE PTVPYNYPP EISKREYVD SIKNDKSGM
101 HKNLAKTSE YTRYKSDHG FESAEMLAT IANTKDIP DLTIEHPDH
151 KEMKOYSIV RVGSTTPE IITIGSHDS INLLPSIMA APGADNNGS
201 TVTNMEALRL YTENFLKRGF RPNNTVEHF YSAEBSGLG SLDFITAYAK
251 OKHVRAMLQ QDMTGVSDE EDEHGIYTD YTPALTDI KLINSLISI
301 PYRDTCCGYA CSDHGSATRN GFGSEVIES EFKINKYIH STMDTLDRLS
351 LAHMAERTKI VLGVIIEIGS WSAW

!!AA_SEQUENCE 1.0
P1:S50587 - hypothetical protein YER084w - yeast (Saccharomyces cerevisiae)
C/Species: Saccharomyces cerevisiae
C/Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 21-Nov-1997
R/Dietrich, F.S.
Submitted to the EMBL Data Library, December 1994
A/Description: The sequence of S. cerevisiae cosmid 9747, 8198, 9781, and
lambda clones 3612 and 6052.
A/Reference number: S50428
A/Accession: S50587
A/Molecule type: DNA
A/Residues: 1-128 <DIE>
A/Cross-references: EMBL:U18839; NID:9603313; PID:9603322; MIPS:YER084w
C/Genetics:
A/Map position: 5R

S50587 Length: 128 February 11, 2000 15:52 Type: P Check: 5798
1 MEELICTPY HSNLEMFLL FFCPSKRAAR GHPEFLETLG YKSNHLIKL
51 LPPSLFTKRV MLNDSSHPPS PDFPTGSSAS PRVKLRPSTL WAPPLTVSSD
101 FAASSSTAP VVTDKPVT AVSKRYQP

!!AA_SEQUENCE 1.0
P1:S64117 - hypothetical protein YGL109w - yeast (Saccharomyces cerevisiae)
N/Alternate names: hypothetical protein G3065
C/Species: Saccharomyces cerevisiae
C/Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 12-Dec-1997
R/Castagnoli, L.; Paoluzi, S.; Minenkova, O.
Submitted to the Protein Sequence Database, May 1996
A/Reference number: S64112
A/Accession: S64117
A/Molecule type: DNA
A/Residues: 1-107 <CAS>
A/Cross-references: EMBL:Z72631; NID:91322653; PID:e243965; PID:91322654;
MIPS:YGL109w
A/Experimental source: strain S288C
R/Castagnoli, L.; Paoluzi, S.; Minenkova, O.
Submitted to the EMBL Data Library, April 1996
A/Reference number: S69421
A/Accession: S69421
A/Molecule type: DNA
A/Residues: 1-107 <CAW>
A/Cross-references: EMBL:X97644; NID:91310710; PID:e240104; PID:91310715
C/Genetics:
A/Map position: 7L

S64117 Length: 107 February 11, 2000 15:52 Type: P Check: 6897
1 MAQNPIDDI QYKRYAKR RMEGCKNSC TIATISLQY YCRSLSHKS
51 CFPFSQNAF SRPLDSESY ETWALELAF CLTRPYCTFH SLEISSQQL

101 TLRLPLG

!!AA_SEQUENCE 1.0
P1:S46801 - hypothetical protein YK1027w homolog - yeast (Saccharomyces cerevisiae)
N/Alternate names: hypothetical protein YHR003c
C/Species: Saccharomyces cerevisiae
C/Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 12-Dec-1997
R/Favell, T.
Submitted to the EMBL Data Library, June 1994
A/Description: The sequence of S. cerevisiae cosmid 9780.
A/Reference number: S46794
A/Accession: S46801
A/Molecule type: DNA
A/Residues: 1-429 <FAV>
A/Cross-references: EMBL:U10555; NID:9500813; PID:9500821; MIPS:YHR003c
C/Genetics:
A/Map position: 8R

S46801 Length: 429 February 11, 2000 15:52 Type: P Check: 5569
1 MANNTWKLIA TTALISVST QLAQSWKREY KLSQAANKRK TVSRPRQYDD
51 HLFREOLARN YAFIEGEMR KIKEQYIVY GAGEVGSWC TMLIRSGCQK
101 IMIDPENIS IDSLSNTHCA VLSDIGKPKV QCLKEHLSKI APWEIKARA
151 KAMIKENSHD LIPADESPT FIVDCIDNLE SKVDLEVAH HNKIDVISM
201 GVATKSDPTR VSINDISMTF FDPISRCVRR KLRKRGATG ISVFSNEM
251 DRRDDILSP IDCEHRAINA VRDEALRHP ELGTMGIFG LSITATLTK
301 VSGYPMKENE VKNRLKTFYDS ILETFOKQMA RLENKESSS LIGLEEVGYI
351 VEEMFRGKSP ISGYSTKLAL TKWEANKEIS LNNVLMTRK EOEIHEKRIL
401 LDGEKLTAVY SEEVLDFTER LFKEEEYYS

!!AA_SEQUENCE 1.0
P1:S64829 - hypothetical protein YLR007w - yeast (Saccharomyces cerevisiae)
N/Alternate names: hypothetical protein L1543
C/Species: Saccharomyces cerevisiae
C/Date: 01-Aug-1995 #sequence_revision 24-May-1996 #text_change 14-Nov-1997
R/Vandenbol, M.; Portetelle, D.; Hilger, F.
Submitted to the Protein Sequence Database, May 1996
A/Reference number: S64742
A/Accession: S64829
A/Molecule type: DNA
A/Residues: 1-336 <VAN>
A/Cross-references: EMBL:Z73179; NID:91360297; PID:e245489; PID:91360298;
MIPS:YLR007w
A/Experimental source: strain S288C
C/Genetics:
A/Map position: 12R

S64829 Length: 336 February 11, 2000 15:52 Type: P Check: 7470
1 MEYHEOVSA PVTGATARKY LLOYLSARG ICHENKLILA LMRLETDAST
51 LNTFWSIQW VDKLNDYINA INVKLNLGK KIRINHGIG RNAVTLAKQO
101 NFESEDNIA IRAHNDYAV LOSIVLPESN RFPVYNLAS TEEKLATRF
151 NONELIEFKW AIDOFMISGE TIVGPALET SIIVEVANI LYATGDSNL
201 AKWRFTSTT VGSTNLFOQO ELTATDIEDL LRLCELMKF YRTOEGFEGI
251 DLRCAIEEE YLTSWNLNT CONCHKLAIO GYACNENECR ENEDETEGNS
301 LSOIHWDCF KHYITHVSKN CDRGSSLIT EGYVYI
```

```
11AA_SEQUENCE 1.0
P1:568481 - hypothetical protein YLR165c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein 19632.11
C:Species: Saccharomyces cerevisiae
C>Date: 20-Jul-1996 #sequence_revision 23-Aug-1996 #text_change 12-Dec-1997
C:Accession: S68481
R:Vardoli M.
Submitted to the EMBL Data Library, July 1996
A:Description: The sequence of S. cerevisiae cosmid 9362.
A:Reference number: S68471
A:Accession: S68481
A:Molecule type: DNA
A:Residues: 1-254 <YNU>
A:Cross-references: EMBL:U01921; NID:g1234842; PID:g1234853; MIPS:YLR165c
C:Genetics:
A:Map position: 12R

568481 Length: 254 February 11, 2000 15:52 Type: P Check: 4630

1 MSLKQIPII PENTHYFIYV KPGIPSQP DCRTGRTHP NIDPTPLER
51 FPAIYSHRE VELCRIVYHL DHCYVGMLI AKTRDGSYVF SFLQKGGNN
101 GKRLQKRYVA IVSSSGRFNK PNNEYIKYGP KYNFLISHGG REITFKREVYD
151 ENCIYQVLVT GKHKQIKNHV SOLIQLPILN DKRHSYVNF PALFNDQIAL
201 HSACITIKIG LQTKHLIPM EHNNTQQLMS RKYVNEGEF TPIKEVILE
251 NMDO

11AA_SEQUENCE 1.0
P1:569848 - hypothetical protein YML101c-a - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C>Date: 23-Aug-1996 #sequence_revision 06-Sep-1996 #text_change 21-Nov-1997
C:Accession: S69848
R:Barrell, B.G.
Submitted to the EMBL Data Library, August 1994
A:Reference number: S47445
A:Accession: S69848
A:Molecule type: DNA
A:Residues: 1-105 <BAR>
A:Cross-references: EMBL:X80835; MIPS:YML101c-a
C:Genetics:
A:Map position: 13L

569848 Length: 105 February 11, 2000 15:52 Type: P Check: 7663

1 MYRIKIKNH KSTGVGCMRL FLTLILLOGI FRTGSMFTIP PASRWFLAAT
51 ALSSVSSGSA CIAGSILDEP NCVSYIDTY EONPSYDEIS RVEPSSDHAK
101 SYMGE

11AA_SEQUENCE 1.0
P1:577699 - inner cell wall mannoprotein ICWP - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YLR390w-a
C:Species: Saccharomyces cerevisiae
C>Date: 13-Jan-1998 #sequence_revision 09-May-1997 #text_change 07-May-1999
C:Accession: S77699; S78080
R:Moukadiri, I.; Armero, J.; Abad, A.; Santandreu, R.; Zueco, J.
J. Bacteriol. 179, 2154-2162, 1997
A:Title: Identification of a mannoprotein present in the inner layer of the
cell wall of Saccharomyces cerevisiae
A:Reference number: S77699; MUID:97234625
A:Accession: S77699
A:Molecule type: DNA
A:Residues: 1-238 <MOD>
R:Du, Z.
Submitted to the EMBL Data Library, January 1995
A:Description: The sequence of S. cerevisiae cosmid 8084.
A:Reference number: S55944

A:Accession: S78080
A:Molecule type: DNA
A:Residues: 1-238 <DUZ>
A:Cross-references: EMBL:U019729; NID:g625097; PID:g2580460; MIPS:YLR390w-a
C:Genetics:
A:Gene: ICWP
A:Cross-references: MIPS:YLR390w-a
A:Map position: 12R
C:Keywords: cell wall

577699 Length: 238 February 11, 2000 15:52 Type: P Check: 7370

1 MRATLLSSV VSLALISKEY LATPPACILA CVAQGVKSSS TCDSLNOVTC
51 YCEHENSARV KCLDISCPNN DADAIVSARK SSCSQONASL GDSSSSASSS
101 ASSSSKASSS TRASSSSASS STRASSSSAS SSTKASSSSA APSSKASSST
151 ESSSSSSSST KAPSSSESSS TYVSSSKOAS STSEAHSSSA ASSTVGOETV
201 SSALPTSTAV ISTFSESGCN VLEAGKSVFT AAVAAALI

11AA_SEQUENCE 1.0
P1:A40550 - peroxisomal membrane protein PAS3 - yeast (Saccharomyces cerevisiae)
N:Alternate names: PAS3 protein; protein YDR329c
C:Species: Saccharomyces cerevisiae
C>Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 05-Dec-1997
C:Accession: A40550; S59794
R:Schfeld, J.; Veenhuis, M.; Kunau, W.H.
J. Cell Biol. 114, 1167-1178, 1991
A:Title: PAS3, a Saccharomyces cerevisiae gene encoding a peroxisomal integral
membrane protein essential for peroxisome biogenesis.
A:Reference number: A40550; MUID:91373453
A:Accession: A40550
A:Molecule type: DNA
A:Residues: 1-441 <HOE>
A:Cross-references: GB:X58407; NID:g41102; PID:g4103
R:Du, Z.
Submitted to the EMBL Data Library, July 1995
A:Description: The sequence of S. cerevisiae cosmid 9798.
A:Reference number: S59783
A:Accession: S59794
A:Molecule type: DNA
A:Residues: 1-441 <DUZ>
A:Cross-references: EMBL:U32517; NID:g914989; PID:g915004; MIPS:YDR329c
A:Experimental source: strain S288C (AB972)
C:Genetics:
A:Gene: SGD:PEX3; PAS3
A:Cross-references: MIPS:YDR329c; SGD:S0002737
A:Map position: 4R
C:Keywords: peroxisome; transmembrane protein
F:23-39/Domain: transmembrane #status predicted <TM1>
F:136-152/Domain: transmembrane #status predicted <TM2>

A40550 Length: 441 February 11, 2000 15:52 Type: P Check: 166

1 MAPNORSRL LQRRGKVL I SLTGIALFT TGSVVVFYK RWLYKQOLRI
51 TEQHRIKQI KRFEQTOED SLTYITELP VWRVNLNEND LNDLSITOL
101 KQKQQLLRA KSSSESSSP LKSKAEIWE LELKSLIKIV TVYTVSSLI
151 LLTRQLNIL TRNEYLDSAI KLIMQOENCN KIONRFYNNV TSWMSDEPK
201 ADDAWYMAK KSKKEGOEY INECAFSLS WMLINKGWS YNEITINOSS
251 IEDGHPHD TLTEEFFSR LTNIFRNYS QIQOONNNL TSILPKDSS
301 GQEFLLSQL DADALTSFHS NTLVNOAVN ELTQCIESTA TSIVLESILN
351 ESHFIMNV GIKTIARKP GOEDQOQOV AVFAMSKDC COEMLQTTAG
401 SSGSGVNEY LATDSVQPL DDLASAVSN FGVSSFSFVK P
```

```

!!AA_SEQUENCE 1.0
P1:S67699 - probable membrane protein YDL151C - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein D1554
C:Species: Saccharomyces cerevisiae
C>Date: 12-Jul-1996 #sequence, revision 12-Jul-1996 #text_change 14-Nov-1997
R:Perez, J.; Blugon, C.; Delaveau, T.; Jacq, C.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S67693
A:Accession: S67693
A:Molecule type: DNA
A:Residues: 1-193 <PDB>
A:Cross-references: EMBL:Z74199; NID:g1431236; PID:e253062; PID:g1431237;
MIPS:YDL151C
A:Experimental source: strain S288C
C:Genetics:
A:Map position: 4L
C:Keywords: transmembrane protein
F:114-130/Domain: transmembrane #status predicted <TM>
S67699 Length: 193 February 11, 2000 15:52 Type: P Check: 4697

1 MSTSLFSLSS PSSSSSMRLR ASNSFPLNF FLDDATPSL SSSSASFSL
51 SAPLSIVSRP FCTRDDPLPS DLNPLKSP FSLKRPFRAN GPLEITCVLF
101 KYLAIRLCWP PAPVTLLEFL LKCFSLPL DSSSFTLDA AASLSLDPL
151 AFAFGINENE GLADPPPLEE ESFNGKRF PLLILLNEC YPA

!!AA_SEQUENCE 1.0
P1:S60916 - probable membrane protein YNL266w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein N0800
C:Species: Saccharomyces cerevisiae
C>Date: 15-Feb-1996 #sequence, revision 01-Mar-1996 #text_change 17-Mar-1999
R:Sen-Gupta, M.; Lyck, R.; Niedenthal, R.K.; Fleig, U.N.; Hegemann, J.M.
submitted to the EMBL data library, October 1995
A:Description: The sequence of a 24152 bp segment from the left arm of
chromosome XIV from Saccharomyces cerevisiae between the BNI1 and the POL2
genes.
A:Reference number: S60909
A:Accession: S60916
A:Molecule type: DNA
A:Residues: 1-139 <SEN>
A:Cross-references: EMBL:X92494; NID:g1045236; PID:g1045244
R:Sen-Gupta, M.; Lyck, R.; Niedenthal, R.K.; Fleig, U.N.; Hegemann, J.H.
submitted to the Protein Sequence Database, April 1996
A:Reference number: S63235
A:Accession: S63235
A:Molecule type: DNA
A:Residues: 1-139 <SEN>
A:Cross-references: EMBL:Z71542; NID:g1302322; PID:e295650; PID:g1302323;
MIPS:YNL266w
A:Experimental source: strain S288C
R:Sen-Gupta, M.; Lyck, R.; Fleig, U.; Niedenthal, R.K.; Hegemann, J.H.
A:Title: The sequence of a 24 152 bp segment from the left arm of chromosome
XIV from Saccharomyces cerevisiae between the BNI1 and the POL2 genes.
A:Reference number: S65111; MUID:96310631
A:Accession: S65118
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-139 <SEF>
A:Cross-references: EMBL:X92494; NID:g1045236; PID:g1045244
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October
1995
C:Genetics:
A:Map position: 14L
A:Note: YNL266w
C:Keywords: transmembrane protein
F:34-50/Domain: transmembrane #status predicted <TM>

F:64-80/Domain: transmembrane #status predicted <TM>
S60916 Length: 139 February 11, 2000 15:52 Type: P Check: 7226

1 MWLINHTYKL LSYFLRKASN RFNSSSSSF SCSEVLELV VFPSCFESI
51 TSFLISFGIL SEFLFSLFC LGFLVYIGL ASALSLSLS KAKIGFSSSL
101 SSISPEGLK SEMMEDDED KERSSLYET SYVAISFK

!!AA_SEQUENCE 1.0
P1:S66918 - probable membrane protein YOR044w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein O2771
C:Species: Saccharomyces cerevisiae
C>Date: 12-Jul-1996 #sequence, revision 12-Jul-1996 #text_change 14-Nov-1997
R:Landt, O.; Hiesel, R.; Unseld, M.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S66907
A:Accession: S66918
A:Molecule type: DNA
A:Residues: 1-157 <LAN>
A:Cross-references: EMBL:Z74952; NID:g1420170; PID:e251967; PID:g1420171;
MIPS:YOR044w
A:Experimental source: strain S288C
C:Genetics:
A:Map position: 15R
C:Keywords: transmembrane protein
F:4-20/Domain: transmembrane #status predicted <TM>
F:24-40/Domain: transmembrane #status predicted <TM>
S66918 Length: 157 February 11, 2000 15:52 Type: P Check: 8726

1 MIELELYVL LVIGSLGYIC RTCIAPLIP FLGIYAFDL LYYRNILVL
51 SQENFYRKLL GRSKTNRPB SPLRLHYSS GDCHDTLIGL VDLRLVFLIS
101 TISHSKRFF STRFQTKSGI NSALDANDVE TTSVSSFTN LHLTSSERG
151 YVINGSI

!!AA_SEQUENCE 1.0
P1:J0589 - pectinesterase (EC 3.1.1.11) precursor - Aspergillus niger
N:Alternate names: pectin methyl-esterase
C:Species: Aspergillus niger
C>Date: 31-Mar-1992 #sequence, revision 31-Mar-1992 #text_change 13-Sep-1998
R:Khanh, N.Q.; Rutkowski, E.; Leidinger, K.; Albrecht, H.; Gottschalk, M.
Gene 106, 71-77, 1991
A:Title: Characterization and expression of a genomic pectin methyl
esterase-encoding gene in Aspergillus niger.
A:Reference number: J0589; MUID:92039066
A:Accession: J0589
A:Molecule type: DNA
A:Residues: 1-331 <KHL>
A:Cross-references: EMBL:X54145; NID:g23372; PID:g23373
A:Experimental source: strain RH5344
A:Note: the authors translated the codon GAG for residue 148 as Ile
R:Khanh, N.Q.; Albrecht, H.; Rutkowski, E.; Loeffler, F.; Gottschalk, M.;
Jany, K.D.
Nucleic Acids Res. 18, 4262, 1990
A:Title: Nucleotide and derived amino acid sequence of a pectinesterase cDNA
isolated from Aspergillus niger strain RH 5344.
A:Reference number: S10487; MUID:90332436
A:Accession: S10487
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-331 <KHZ>
A:Cross-references: EMBL:X52902; NID:g23368; PID:g23369
R:Markovic, O.; Joernvall, H.
Protein Seq. Data Anal. 3, 513-515, 1990
A:Title: Tomato and Aspergillus niger pectinesterases. Correlation of
differences in existing reports: large species variations.

```

A:Reference number: S13959; MID:91212339  
 A:Accession: S13959  
 A:Molecule type: protein  
 A:Residues: 18,'XG','21-32','F','34 <MAR>  
 A:Experimental source: strain 8 MX-41  
 C:Comment: This enzyme belongs to the group of pectic enzymes and catalyzes the hydrolysis of methyl ester groups of galactaronic acid residues of pectin.  
 C:Genetics:  
 A:Gene: pmeA  
 A:Introns: 39/3; 85/2; 112/1; 161/1; 181/2; 246/1  
 C:Keywords: carboxylic ester hydrolase; glycoprotein; hydrolase  
 F:1-17/Domain: signal sequence #status predicted <Sig>  
 F:18-33/Product: pectinesterase #status experimental <Mat>  
 F:95,283,302/Binding site: carbohydrate (asn) (covalent) #status predicted  
 J0589 Length: 331 February 11, 2000 15:52 Type: P Check: 609 ..

1 MKKSLASVL FAATALAASR MTPASGAIV AKSGGDYDTI SAAVDALSTT  
 51 STETOTIFIE EGSYDEQVYI PALSGKLIVY GQEDITTYT SMLVITTHAI  
 101 ALADVNDDE TATLRNTAEG SAIYNLINIAN TCGAACHQAL AVSAYASEOG  
 151 YVACQFTGY DTLAETGYQ VYAGTYIEGA VDFIFGQAR AMFHECDIRY  
 201 LEPSSASIT ANGRSESD SYVIRKSTV AAAGNDVSS GTYLIGRPMS  
 251 QYARVCFORT SMTDVINHLG WTEMSTSTPN TENVTEVYG NTGTGAEGR  
 301 ANFSSEITEP ITISWLGSD WEDWDTSYI N

11AA\_SEQUENCE 1.0  
 P1:S03833 - hypothetical protein 1 - chestnut blight fungus  
 C:Species: Cryphonectria parasitica, Endothia parasitica (chestnut blight fungus)  
 C:Date: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 09-Sep-1997  
 C:Accession: S03833  
 R:Rae, B.P.; Hillman, B.L.; Tartaglia, J.; Nuss, D.L.  
 EMBL J. 8, 657-663, 1989  
 A:Title: Characterization of double-stranded RNA genetic elements associated with biological control of chestnut blight: organization of terminal domains and identification of gene products.  
 A:Reference number: S03833; MID:89251594  
 A:Accession: S03833  
 A:Molecule type: DNA  
 A:Residues: 1-319 <RAE>  
 A:Cross-references: EMBL:X14524; NID:92624; PID:92625  
 A>Note: the authors translated the codon CAG for residue 156 as Gly  
 S03833 Length: 319 February 11, 2000 15:52 Type: P Check: 3269 ..

1 MAQLRKPQS LVLSSEVDPT TVDPFVSVRT EEVVPACIT LMERYDSGCD  
 51 VGPPLSHDL RRLRTPDVC KCOVHELEPT VLKSGSTGV PEHPAVLAF  
 101 IGRPRCSLE QRTKELDSRF LQLVHGGLPA RPSYMIARP RPYRGICSSR  
 151 NSLSAQFGG YCYLSAIVDS ARWRYARTTG WCVNAVLYR LIOWGRNRF  
 201 GSNQIEKAV DHVYHVAVDA EYQSEDDGAL FYQAILGLAE KQPLARIGR  
 251 LNLIAEFAP GSALREPEPT PQVTRRKSTI EMTGRDPTV FRWQNGYGH  
 301 QHPCSCGYG VEROFRPS

11AA\_SEQUENCE 1.0  
 P1:A34051 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) 29/21K chain precursor  
 N:Neurospora crassa  
 N:Alternate names: complex I dehydrogenase 29/21K chain  
 C:Species: Neurospora crassa  
 C:Date: 07-Jun-1990 #sequence\_revision 07-Jun-1990 #text\_change 09-Sep-1997  
 C:Accession: A34051  
 R:Videla, A.; Tropchug, M.; Werner, S.

Biochem. Biophys. Res. Commun. 166, 280-285, 1990  
 A:Title: Primary structure, in vitro expression and import into mitochondria of a 29/21-KDa subunit of complex I from Neurospora crassa.  
 A:Reference number: A34051; MID:90147712  
 A:Accession: A34051  
 A:Molecule type: mRNA  
 A:Residues: 1-201 <YAD>  
 A:Cross-references: GB:M32244; NID:9168778; PID:9168779  
 C:Keywords: mitochondrion; NAD; oxidoreductase  
 A34051 Length: 201 February 11, 2000 15:52 Type: P Check: 7750 ..

1 MAAKYVTVG KTAGGVVPY SOKTYVOSYG VMERIRARA IDPNRNSGVP  
 51 LVPYRNRPSP GSIDPLAYDD PVTIAGIA DNYWKRDR RNPPLSYVG  
 101 QAEVALLSV GSATHPRVEL VGENSKQLY AADACKTGS LAKYEGTGV  
 151 EAGKVLAEI GGLPPLPGE KLGEGKWDV YKQLAEPFS YEAVPCRSF  
 201 S

11AA\_SEQUENCE 1.0  
 P1:S16556 - hypothetical protein 1 - Podospira anserina  
 C:Species: Podospira anserina  
 C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 24-Jul-1997  
 C:Accession: S16556  
 R:Turcq, B.; Delau, C.; Denayrolles, M.; Begueret, J.  
 Mol. Gen. Genet. 228, 265-269, 1991  
 A:Title: Two allelic genes responsible for vegetative incompatibility in the fungus Podospira anserina are not essential for cell viability.  
 A:Reference number: S16556; MID:91360075  
 A:Accession: S16556  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-289 <TUR>  
 S16556 Length: 289 February 11, 2000 15:52 Type: P Check: 4238 ..

1 MSEFGLVAG ALNVAGLFNN CVDFEYVOL GRPFGRDYER COLRLDIARA  
 51 RLKRWGAIVK INDDPRFSS APTKSVOLA KSIVEIILS FESAOKTSKR  
 101 YELVADQDL VYFEDKMKP IGRALHRLN DLYSRROKOT SLAKTAMAL  
 151 YDGSLEKIV DOYARVDEL EKAPIEAVC HKLAIEIDE VEDASLTIL  
 201 KDAAGCIDAA MSDAAQKID AIVGRNSAKD IRTREERAVQ LGNVYTAAL  
 251 HGGIRISDOT TNSVETVYVK GESRVILCNE YGKKGWDM

11AA\_SEQUENCE 1.0  
 P1:S16557 - hypothetical protein 2 - Podospira anserina  
 C:Species: Podospira anserina  
 C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 24-Jul-1997  
 C:Accession: S16557  
 R:Turcq, B.; Delau, C.; Denayrolles, M.; Begueret, J.  
 Mol. Gen. Genet. 228, 265-269, 1991  
 A:Title: Two allelic genes responsible for vegetative incompatibility in the fungus Podospira anserina are not essential for cell viability.  
 A:Reference number: S16557; MID:91360075  
 A:Accession: S16557  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-289 <TUR>  
 S16557 Length: 289 February 11, 2000 15:52 Type: P Check: 4187 ..

1 MSEFELVAG ALVAGLFNN CVACFEYVOL GRHFGDXYER COLRLDIARY  
 51 RLKRWGEAVQ INDDPRFSS APIKSVOLA KSIVEIILS FESAOKTSKR  
 101 YELVADQDL VYFEDKMKP IGRALHRLK DLYSRROKOT SLAKTAMAL

Mon Feb 14 08:07:32 2000

pir.cat

Page 170

151 YGKSLKELIV DOVAGFVDEL EKAFPIEAVC HKLAENEIEE VEDEASLTIL  
201 KDAAGGIDAA MSDAAAKRID AIYGRNSAKD IREKERRARQ LGNYTARL  
251 HGEIRISDOT TNSYETVWGR GESKVLIGNE YGKRGFMDN

!!AA\_SEQUENCE 1.0  
P1:JN0451 - phosphoribosylanthranilate isomerase (EC 5.3.1.24) - fungus  
N:Alternate names: PRA1  
C:Species: Filobasidiella neoformans, Cryptococcus neoformans  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 10-Oct-1997  
R:Perfect, J.R.; Rude, T.H.; Penning, L.M.; Johnson, S.A.  
A:Title: Cloning the Cryptococcus neoformans TRP1 gene by complementation in  
Saccharomyces cerevisiae  
A:Accession: JN0451; MID:93083987  
A:Residues: 1-312 <PER>  
A:Residues: 1-312 <PER>  
A:Cross-references: GB:M74901; MID:9167395; PID:9167396  
C:Genetics:  
A:Gene: TRP1  
C:Superfamily: trpF homology  
C:Keywords: Intramolecular oxidoreductase; Isomerase  
F:77-312/Domain: trpF homology <TRF>

JN0451 Length: 312 February 11, 2000 15:52 Type: P Check: 2342

1 MSTSAYVNA LNRDQVYCA LSGISSHEDV EKYKEGVKG VYGEALMRA  
51 SOTKATLRL IGLPPELVYS KPRPLVYKIG IRSTNDAKLA INAGADLIGV  
101 ILVPGTKRCI STTAREISA LVQASARSQSS SKPLEPSSLS PMFTQSALL  
151 SSRKRPVLVG VFONQSLSDI LSAVDEIGLD LVQIHGDEPO AVAKFIPIPV  
201 VAVFVSEEG IVRGEIRRP GLNQAILLDA GGAAGGGGEG KAFPEHAKR  
251 LIQSGEVSE GHVPLPYILA GGLTPEVNGQ ALNRLVKARG VMASVSGSE  
301 REGSRRLRL RS

!!AA\_SEQUENCE 1.0  
P1:B4418 - surface antigen - Trypanosoma brucei  
C:Species: Trypanosoma brucei  
C:Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 07-Feb-1997  
R:Accession: B4418  
J:Bio1. Chem. 268, 1894-1900, 1993  
A:Title: Cloning of a novel surface antigen from the insect stages of  
Trypanosoma brucei by expression in COS cells.  
A:Reference number: A4418; MID:9313193  
A:Contents: 427 Nucleotides  
A:Accession: B4418  
A:Status: preliminary  
A:Molecule type: nucleic acid  
A:Residues: 1-412 <TAC>  
A:Note: sequence inconsistent with the nucleotide translation  
C:Keywords: surface antigen

B4418 Length: 412 February 11, 2000 15:52 Type: P Check: 3830

1 MCIEOLVHSD GEFRTAVAC LCCLLIGSPV LIGGVLFSL SDPPRDNFRK  
51 AVSAPRPKL ESWTGTSDV KATVRROSLS VAGLSIPSY YTEATPVASG  
101 NNDGSLVVK VNITVAPFT RRSPLHATRE RMFSCSSSQC SGYSRCDQ  
151 EKHEDFRNKC YSEGGYSTQ SSKRLGKGC GYCKOEYVLS KLYLVAASDG

201 KGEYESTQY QSALYSFGL SQYEAVPOD KYOVOLYSEB DPTALERET  
251 MGEFGEVFN RMGIACIYA GSLLILLEIA VCVVVCFCCL KRGSSSNDT  
301 GDPPTQGDG SPYTYGQSQP PPPRGYAYGQ PLPQCGTYG QPPPOGAT  
351 YGQPPPPQG YPYGQPPPPQ QGGTYTKPP PQGGTYGQP PPPPOGHSY  
401 GQACPPPNP TV

!!AA\_SEQUENCE 1.0  
P1:T14622 - hypothetical protein - Trypanosoma cruzi  
C:Species: Trypanosoma cruzi  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
R:Anderson, B.; Aslund, L.; Petersson, U.  
A:Submitted to the EMBL Data Library, March 1998  
A:Description: 93.4 Kd of complete sequence from chromosome 3 of Trypanosoma  
cruzi  
A:Reference number: 218159  
A:Accession: T14622  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-425 <AND>  
A:Cross-references: EMBL:AF052833; MID:93063554; PID:93063555; PIDN:AA014090.1  
C:Genetics:  
A:Map position: 3

T14622 Length: 425 February 11, 2000 15:52 Type: P Check: 9305

1 MSRLNPATLE SIMOVGLVG RSPFPSSSG KARESGRIS TSDRLRLKA  
51 VIEEDALSTI VGGMIIPFLV VEKTTGLR RMIAWRDKN RDDYEANAP  
101 LSHISHYLP VMAEASCID LKTSFQVSL PRTRHLFC RVEGTVEL  
151 TRLPGYKAS PEILOITSA IAGTVVHR LCAPPPIVI VWIGNIRIT  
201 GSKGATIME AQLRNPDGC HASMGEEES GATOTFLMV RPDHTRAVS  
251 LSKFVSVR AMPALNSSTI AEMEVSAR LYAAAILGTR SCQYFFIKA  
301 VRRRLALNR GIVQETSPN VPPAVGLG RLRLIENIT VSESSPRKR  
351 HRPSSSMHR SMDGEPLFQ TPATIKLPEE NGRNSLFLSC RPRRAVLAL  
401 SAFRRLAH HSLGGQHPA ARSGE

!!AA\_SEQUENCE 1.0  
P1:T15458 - hypothetical protein C08A9.7 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
R:Accession: T15458  
Submitted to the EMBL Data Library, December 1995  
A:Description: The sequence of C. elegans cosmid C08A9.  
A:Reference number: 218353  
A:Accession: T15458  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-331 <LAT>  
A:Cross-references: EMBL:U42844; MID:91125797; PID:91125804; PIDN:AA053819.1;  
GSPDB:G000028; CESP:C08A9.7  
A:Experimental source: strain Bristol N2; clone C08A9  
C:Genetics:  
A:Gene: CESP:C08A9.7  
A:Map position: 10  
A:Introns: 100/1; 188/2; 247/3

T15458 Length: 331 February 11, 2000 15:52 Type: P Check: 2648

1 MVSATRVRR SSTTSATQ ORTPSLMPA SPFITMDEYL EKENRETVN

```

51 ASKDIANKRL ALTLLEIPE MMRGPGMVA KKKOAFGAM YRRTGKIYRC
P1:T15482 - hypothetical protein C10A4.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15482
R:Pauley, A. to the EMBL Data Library, March 1995
A:Description: The sequence of C. elegans cosmid C10A4.
A:Reference number: Z18358
A:Accession: T15482
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-217 <PDB>
A:Cross-references: EMBL:U23454; NID:g733580; PID:g733583; PIDN:AAC46517.1;
CESP:C10A4.3
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:C10A4.3
A:Introns: 31/1; 59/3; 85/2

T15482 Length: 217 February 11, 2000 15:52 Type: P Check: 121 ..

1 MARMIRSFRI ILMCSFISI CLYTAMTE SSINRIIPR TOCLTCNPHQ
51 EEPYQILNHL SQISEFTKIN FAVDLCLFVI TTFNLVKMQG SGADDFLISI
101 GVFCIKYILIP YMLNFFSGA HSPAHNISTYL LNNEVCGTIG NCLGNVARI
151 MYSCISYVII MMLAFPALIA SMGWYIITFV RTIDQPLPMV TEQDDLPAY
201 SVLFPSQOOL LKKEDEC

11AA_SEQUENCE 1.0
P1:T15182 - hypothetical protein C18E3.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15182
R:Connell, M.; Maggi, L.
submitted to the EMBL Data Library, April 1997
A:Description: The sequence of C. elegans cosmid C18E3.
A:Reference number: Z18304
A:Accession: T15182
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-150 <CON>
A:Cross-references: EMBL:AF000265; NID:g1947147; PID:g1947151; PIDN:AAB52944.1;
GSPDB:GN00019; CESP:C18E3.4
A:Experimental source: strain Bristol N2; clone C18E3
C:Genetics:
A:Gene: CESP:C18E3.4
A:Map position: 1
A:Introns: 33/3; 104/2; 123/3

T15182 Length: 150 February 11, 2000 15:52 Type: P Check: 9118 ..

1 MSLTMDKKS RMKRIYVVT TTTLDLELY YGAEVLSVY VSVVSIQNTQ
51 QFVVYDIKIAL LYINCNDLIP KASKLYITIIQ TILPLSRNLL IROPKILTF
101 ISOINVDKOC NSPMFLKTF SPSSKLLKFIK NFVQGMNMTF QKKQIITIH

51 ASKDIANKRL ALTLLEIPE MMRGPGMVA KKKOAFGAM YRRTGKIYRC
P1:T15715 - hypothetical protein C30G12.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15715
R:Latreille, P.
submitted to the EMBL Data Library, July 1995
A:Description: The sequence of C. elegans cosmid C30G12.
A:Reference number: Z18393
A:Accession: T15715
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-280 <LAT>
A:Cross-references: EMBL:U21319; NID:g687832; PID:g687834; PIDN:AAC46671.1;
CESP:C30G12.2
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:C30G12.2
A:Introns: 184/3; 199/3

T15715 Length: 280 February 11, 2000 15:52 Type: P Check: 4709 ..

1 MADVADPEVE SRPQRFPEKNI MITGANRGIG FGLYKHFLY DCIELLIATC
51 RNPEKADELN ALKNDRLHY IALNVDDDS IKKYFDEVSS LVSSGLMML
101 INNAGILLYE EVDGPKICRK TMKQLEINS VSAVILTOIF DPLIKTAASA
151 AEGDEASIDR ASINISSTM ASIENNNGCF DGPKRFFHY FLVNSOEMTT
201 AYMSKSALN AFAROSEMEL SKYHILVTSF CPGWRTDMG GDNADLDVNE
251 STKLSANIL RLDNNNGLY FDRFLHIPN

11AA_SEQUENCE 1.0
P1:T15729 - hypothetical protein C31H1.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15729
R:Le, T.
submitted to the EMBL Data Library, December 1995
A:Description: The sequence of C. elegans cosmid C31H1.
A:Reference number: Z18395
A:Accession: T15729
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-304 <LET>
A:Cross-references: EMBL:U42848; NID:g1125817; PID:g1125822; PIDN:AA63610.1;
CESP:C31H1.6
C:Genetics:
A:Gene: CESP:C31H1.6
A:Introns: 5/1; 38/3; 77/3; 106/3; 215/3; 234/1; 271/3

T15729 Length: 304 February 11, 2000 15:52 Type: P Check: 770 ..

1 MSDECSISPO KSPAFSPEDE KMLKSNRANE POSFVKGND AAKSDLIDLEE
51 VYTNNAKTAF DVTKSKAEGK IKKRYLSYNI NMTKLFQEGS EPTYIILGAK
101 PNWVDAKQDE KVIRKTFITN ITOROEYSF KTERCTRSS TVIIEKGYCR
151 GAVYSIKLKT PCEVEANAG FTEVNLNHI GENTSEBELT WGVDSQVAVP
201 PGAEYIAELV IIEDIVTRDF RIENRLSGV LVITINLKEN NSLVITIEGN
251 ICDIINGMPD YTAGFRFDG PAIYETRGO CIFRFEGIEOK VRINEFSLNS
301 SRRY

11AA_SEQUENCE 1.0
P1:T16486 - hypothetical protein F56D3.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans

```





544865 Length: 336 February 11, 2000 15:52 Type: P Check: 399  
 1 MTLLEVRAS TKITITLISR IEMGRILEKP RILGTWSEFR IGLNCESLSG  
 51 AIQRLPNESK LSEFKKIGV GKKRSVTIG VFPSDDTSF QHFOIAEKL  
 101 SKRYVAEFI NENSKPAITT YHAEKORT DYSGKDPAT LMEFITKSI  
 151 PSIIISNGE TTDLEHOKR PLIIIGAG EPEFSLSLA RODAKTYIF  
 201 TKIDSESEMI KTKRKALGV ENESITVPLN KDRVHRIPIS KRKCDHLOK  
 251 ILOMITTSEA DOVLSTKEPH PLRYLOKEV NEVGFEEETL VLPDHTLFLD  
 301 SDFSRHPP1 TEGGGGCGPF MOGGAPTES SQHSEL  
 11AA:SEQUENCE 1.0  
 P1:S06613 - chorion protein sl9 - fruit fly (Drosophila grimshawi)  
 C:Species: Drosophila grimshawi  
 C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 24-Sep-1998  
 C:Accession: S06613  
 R:Kertner-Cruzado, J.C.; Swimmer, C.; Fenerjian, M.G.; Kafatos, F.C.  
 Genes 119, 663-677, 1998  
 A:Title: Evolution of the autosomal chorion locus in Drosophila. I. General  
 organization of the locus and sequence comparisons of genes sl5 and sl9 in  
 evolutionarily distant species.  
 A:Reference number: S06612; MUID:88297142  
 A:Accession: S06613  
 A:Molecule type: DNA  
 A:Residues: 1196 <MAP>  
 A:Cross-references: GB:X53422; EMBL:X12635; NID:97419; PID:97422  
 C:Genetics:  
 A:Gene: sl9-1  
 A:Cross-references: FlyBase:FBgn0012320  
 A:Map position: 5  
 A:Introns: 5/3  
 S06613 Length: 196 February 11, 2000 15:52 Type: P Check: 4664  
 1 NMTFATLAI ISACLAVGC GYGSPICYG GPINVLGRV SSIGQSGDG  
 51 AAASAAASG GDNQPEVETIA GGAPRYGSSO NLRPILLNSG YHGLDNWIG  
 101 RIAQIVGGGR SLGCHLGHL GCHLGRIGG NYGGRYRRR FLYOPAGATL  
 151 LYPGQNSYR ISSPEYSKY ILPRAAPV AKLYIPONTV GSQVY  
 11AA:SEQUENCE 1.0  
 P1:S69241 - Dreg-5 protein - fruit fly (Drosophila melanogaster)  
 C:Species: Drosophila melanogaster  
 C:Date: 04-Dec-1997 #sequence\_revision 12-Dec-1997 #text\_change 16-Apr-1999  
 C:Accession: S69241  
 R:Van Gelder, R.N.; Krasnow, M.A.  
 Embo J. 15, 1625-1631, 1996  
 A:Title: A novel circadianly expressed Drosophila melanogaster gene dependent  
 on the period gene for its rhythmic expression.  
 A:Reference number: S69241; MUID:96203080  
 A:Accession: S69241  
 A:Molecule type: mRNA  
 A:Residues: 1298 <YAN>  
 A:Cross-references: GB:U65105; GB:S81693; NID:91513135; PID:91513136  
 C:Genetics:  
 A:Gene: Dreg-5  
 S69241 Length: 298 February 11, 2000 15:52 Type: P Check: 7903  
 1 MTLAKYILA CCLGAFHIO ISSSAIPW EFLTRNEMK HLYSTFQALV  
 51 SVCKSTRAY GGLPVNOCK NLIGYSAKL OTLSVQLBA LDPORDANE  
 101 LWSISGDM PSGLSVYTR QPLQPLPTP PASLITLIR QQLPHGASHA  
 151 HPIOSSGSAT NPFESEQK HKYAMDMDKA YGYPQSSSE LPVAAALISE  
 201 PSKRFLTCPL VIRVRPDSP VEEDKMPLP RDEDLPYLS WSGRSAGQA  
 251 PODRHNOLK AALRLHPAE RPPATPPDAE ASVPATGVR SRSEDDQA  
 11AA:SEQUENCE 1.0  
 P1:A4254 - ecdysone-induced protein - fruit fly (Drosophila melanogaster)  
 C:Species: Drosophila melanogaster  
 C:Date: 09-Sep-1987 #sequence\_revision 09-Sep-1987 #text\_change 16-Feb-1997  
 C:Accession: A4254  
 R:Cherbas, L.; Schultz, R.A.; Koehler, M.M.D.; Savakis, C.; Cherbas, P.  
 J. Mol. Biol. 189, 617-631, 1986  
 A:Title: Structure of the E1p28/29 gene, an ecdysone-inducible gene from  
 Drosophila.  
 A:Reference number: A4254; MUID:87060956  
 A:Accession: A4254  
 A:Molecule type: DNA  
 A:Residues: 1-255 <CH>  
 C:Genetics:  
 A:Gene: FlyBase:E1p71CD  
 A:Cross-references: FlyBase:FBgn0000565  
 A:Introns: 17/3; 82/3; 169/3  
 C:Keywords: alternative splicing  
 A4254 Length: 255 February 11, 2000 15:52 Type: P Check: 68  
 1 MSLLITSVT HPEIKDLSTV RNEOKELNIS PYHDVNTKA TATFGMGFW  
 51 GASIVGATR GVLRTVGYA GGSSDLPTFR KMGDTEYLE IDYDPAIVSF  
 101 KELLDLFVNN HEYGLTPIK ROYASLIYH DEOKOVANA SKLEODERA  
 151 PEITTEIAS KENFYPAEAY HOKYRLQGRK DLASLNLSP KLIQTSYVAT  
 201 KLNQYLAGV GIEQFKAEAE TTGSDAHPPA VLLPRAER GPGSLILTP  
 251 NVHR  
 11AA:SEQUENCE 1.0  
 F1:S78042 - Ig mu chain C region, membrane-bond form - channel catfish  
 (fragment)  
 C:Species: Ictalurus punctatus (channel catfish)  
 A:Variety: Channel catfish  
 C:Date: 24-Oct-1998 #sequence\_revision 24-Oct-1998 #text\_change 24-Oct-1998  
 C:Accession: S78042  
 R:Wilson, M.R.; Marcuz, A.; van Ginkel, F.; Miller, N.W.; Clem, L.W.;  
 Middleton, D.; Warr, G.W.  
 Nucleic Acids Res. 18, 5227-5233, 1990  
 A:Title: The immunoglobulin M heavy chain constant region gene of the channel  
 catfish, Ictalurus punctatus: an unusual mRNA splice pattern produces the  
 membrane form of the molecule.  
 A:Reference number: S12833; MUID:90384824  
 A:Accession: S78042  
 A:Molecule type: DNA  
 A:Residues: 1-351 <WIL>  
 A:Cross-references: EMBL:X52617  
 A:Experimental source: erythrocyte  
 C:Genetics:  
 A:Introns: 101/3; 203/3; 304/3; 349/3  
 A:Keywords: alternative splicing; glycoprotein; heterotrimer; immunoglobulin;  
 transmembrane protein  
 S78042 Length: 351 February 11, 2000 15:52 Type: P Check: 1000  
 1 VQSAFSLFP VMCGSASDG LVTLGCVTRD LASADGLSFI WPDASSALT  
 51 DVQYPAVQA TGGYVSASH RVKASDMGN KFTCEVKG LSGKASLQK  
 101 PYRELHSL ILTPTQTEI DNGTATFVCL AIPSPKSHT FKWTLEKIDI  
 151 SNKVENIVS QNKGNFALS VLELSASEWT SSTSPKCEF QQXNHVKE

Mon Feb 14 08:07:32 2000

pl.r.cat

Page 174

201 ASYAPBDTKO POKYITGPST EDILIRRAQ LECRAEGDTG FKSIRKLIGN  
251 REISSISNNS SKTVSLQTH IGFEEMINGT EFICEVEHEA FTQOYEKVF  
301 KRENVCIMST EIFHEMEMD DDNMANTALT FVFLLITLF YSIGVTVFVY  
351 K

11AA-SEQUENCE 1.0  
P1:JCS761 - early lactation protein precursor - brush-tailed possum  
C/Species: Trichosurus vulpecula (brush-tailed possum)  
C/Date: 15-Feb-1997 #sequence-revision 13-Mar-1997 #text-change 16-Jul-1999  
C/Accession: JCS761  
R/Plotte, C.P.; Grigor, M.R.  
Arch. Biochem. Biophys. 330, 59-64, 1996  
A/Title: A novel marsupial protein expressed by the mammary gland only during the early lactation and related to the Kunitz proteinase inhibitors.  
A/Reference number: S69288; MUID:96230240  
A/Accession: S69288  
A/Molecule type: mRNA  
A/Residues: 1-102 <PRO>  
A/Cross-references: EMBL:U34208; NID:q1565292; PID:q1002806  
A/Accession: S74274  
A/Molecule type: Protein  
A/Residues: 21-30 <PIIT>  
C/Superfamily: animal Kunitz-type proteinase inhibitor homology  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-102/Product: early lactation protein #status experimental <MAT>  
F:43-93/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>  
S69288 Length: 102 February 11, 2000 15:52 Type: P Check: 5524

1 MKTITATCL ASLVGMSS EKLDIRAN SLENLRVP SLCLPSGRC  
51 NCOSQILRF VYATSHCEV FLYSGGNG NNFDSLECL KTCRLKRYN  
101 NN  
11AA-SEQUENCE 1.0  
P1:TL143 - ATPase subunit 8 - orangutan mitochondrion (SGC1)  
C/Species: mitochondrion Pongo pygmaeus (orangutan)  
C/Date: 20-Sep-1999 #sequence-revision 20-Sep-1999 #text-change 20-Sep-1999  
C/Accession: TL143  
R/Foral, S.; Hayasaka, K.; Kondo, R.; Tsugane, K.; Takahata, N.  
Proc. Natl. Acad. Sci. U.S.A. 92, 532-536, 1995  
A/Title: Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs.  
A/Reference number: 159384; MUID:95132634  
A/Accession: TL143  
A/Status: Preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-68 <HOR>  
A/Cross-references: EMBL:D38115; NID:d1027246; PID:d1007887; PIDN:BA07309.1  
C/Genetics:  
A/Genome: mitochondrion  
A/Genetic code: SGC1  
C/Keywords: mitochondrion  
TL143 Length: 68 February 11, 2000 15:52 Type: P Check: 3550

1 MPQNTTTL TVITPLAL FLITQKLN SHLHPTPK FTKTRPHAK  
51 WELKTKIYS PHSLPQS

11AA-SEQUENCE 1.0  
P1:JCS761 - cytokine-inducible SH2 protein 3 - human  
C/Species: Homo sapiens (man)  
C/Date: 24-Jan-1998 #sequence-revision 13-Mar-1998 #text-change 07-May-1999  
C/Accession: JCS761  
R/Masuhara, M.; Sakamoto, H.; Matsumoto, A.; Suzuki, R.; Yasukawa, H.; Mitsui, K.; Wakitani, T.; Tanimura, S.; Sasaki, A.; Misawa, H.; Yokouchi, M.; Ohtsuda, M.; Yoshimura, A.  
Biochem. Biophys. Res. Commun. 239, 439-446, 1997

A/Title: Cloning and characterization of novel cis family genes.  
A/Reference number: JCS760; MUID:98008857  
A/Accession: JCS761  
A/Status: nucleic acid sequence not shown  
A/Molecule type: mRNA  
A/Residues: 1-225 <MAS>  
A/Cross-references: DDBJ:AB006967; NID:q2463522; PID:d1023405; PID:q2463523  
C/Comment: This protein plays a role in the negative regulation of cytokine signaling by interacting with specific targets.  
F:45-144/Domain: SH2 homology <SH2>  
JCS761 Length: 225 February 11, 2000 15:52 Type: P Check: 447

1 MYTHSKPPAA GMSRPIDSL RKTFFSKSE YOLVNAVVR LQSGFYWSA  
51 VTGEANULL SAEPACTFLI RDSSDORHFF TISVKTOSGT KNLRIQCEG  
101 SFSIQSDPRS TQVPRFDCV LKLVHMYMP QGAPSFSP TEPSSVPEQ  
151 PSAPPLPGSP PRRAVYISG GEXIPVLGR PLSSNVATIQ HLCRTVNGH  
201 LDSYEKVTQL PGPREFIDQ YDAPL

11AA-SEQUENCE 1.0  
P1:S43928 - estradiol 17beta-dehydrogenase (EC 1.1.1.62) type 3 - human  
N/Alternate names: 17-beta-hydroxysteroid dehydrogenase  
C/Species: Homo sapiens (man)  
C/Date: 10-May-1995 #sequence-revision 10-Nov-1995 #text-change 28-May-1999  
C/Accession: S43928  
R/Beisler, W.M.; Davis, D.L.; Wu, L.; Bradshaw, K.D.; Patel, S.; Mendonca, B.B.; Ellison, K.O.; Wilson, J.D.; Russell, D.W.; Andersson, S.  
Nature Genet. 7, 34-39, 1994  
A/Title: Male pseudohermaphroditism caused by mutations of testicular 17-beta-hydroxysteroid dehydrogenase 3.  
A/Reference number: S43928; MUID:94355972  
A/Accession: S43928  
A/Status: Preliminary  
A/Molecule type: mRNA  
A/Residues: 1-310 <GEI>  
A/Cross-references: GB:U05659; NID:q531161; PIDN:AA050066.1; PID:q531162  
C/Genetics:  
A/Gene: GDB:HSD17B3  
A/Cross-references: GDB:347487; OMIM:264300  
A/Map position: 9q22-9q22  
C/Superfamily: short-chain alcohol dehydrogenase homology  
F:49-229/Domain: short-chain alcohol dehydrogenase homology <SADH>  
S43928 Length: 310 February 11, 2000 15:52 Type: P Check: 3434

1 MGDVLEPFI LGLVAVLAC LAKCVRESRC VILNWKVP KSELRSMGW  
51 AVTAGADGI GKAYSFEAK RGLNVVLSR TLEKLEIAT EIBRTGRSV  
101 KIIGADFTK DIYEHKEKL AGLEIGILVN NVGMLPNLP SHLNPADDEI  
151 QSLHCNITS YKMTQILIK HMESROGLI LNISSGIALF PMPLYMYS  
201 SKAFVCAFSS ALQERYAKE VITQVLPYA VSIAMTYLN TNYITKTADE  
251 FVKEISLNYT IGGETCGCLA HELIAGFLSL IPAWAFSGA FORLLITHYV  
301 AYKLNTRYR

11AA-SEQUENCE 1.0  
P1:JEO175 - freezed protein-1b - human  
C/Species: Homo sapiens (man)  
C/Date: 03-Jul-1998 #sequence-revision 10-Jul-1998 #text-change 07-May-1999  
C/Accession: JEO175  
R/Hu, E.; Zhu, Y.; Fredrickson, T.; Barnes, M.; Kelsell, D.; Beeley, L.; Brooks, D.  
Biochem. Biophys. Res. Commun. 247, 287-293, 1998

A:Title: Tissue restricted expression of two human frzbs in preadipocytes and pancreas.  
 A:Reference number: JEO174; MUID:98308108  
 A:Accession: JEO175  
 A:Molecule type: mRNA  
 A:Residues: 1-317 <HUA>  
 A:Genetics:  
 A:Gene: hFRP-1b  
 A:Map position: 5q14.3-q12.1  
 JEO175 Length: 317 February 11, 2000 15:52 Type: P Check: 3945 ..

1 MRAAAAAGV RTAALLILG ALHAPARCE EYHYGMOAE PLHGRTSKP  
 51 PGCDIPADL PLCHTVGYKR MRLPILTHE SLAEVROAS SWLPLAKRC  
 101 HSDTVFGLS LFAPVCLDRP IYPCRSLEA VRAGCAPLME AYGPWPBML  
 151 HCHKEPLDND LCIAVDFGHL PATAPPVTKI CAOCMEHSA DGLMEQMCSS  
 201 DFVYKMRITE IKIENGDRKL IGAOKKKLL KPGPLKEDT KRLVLHMKNG  
 251 AGCPCQQLDS LAGSFLVMGR KYDQGLLMA YRWMDKKNKE MKFAVKFMS  
 301 YPCSLTYPF YGAAPPH

!!AA\_SEQUENCE 1.0  
 FI:177373 - gene N-ras protein - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 28-Feb-1997  
 C:Accession: I77373  
 R:Kiba, T.; Tsuda, H.; Pairojkul, C.; Inoue, S.; Sugimura, T.; Hirohashi, S.  
 Mol. Carcinog. 8, 312-318, 1993  
 A:Title: Mutations of the p53 tumor suppressor gene and the ras gene family in intrahepatic cholangiocellular carcinomas in Japan and Thailand.  
 A:Reference number: I57484; MUID:94107479  
 A:Accession: I77373  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-22 <RES>  
 A:Cross-references: GB:S68581, NID:9544858, PID:9544859  
 A:Genetics:  
 A:Gene: N-ras  
 I77373 Length: 22 February 11, 2000 15:52 Type: P Check: 9167 ..

1 ETCIDITLL LTRRLPLCIC HQ

!!AA\_SEQUENCE 1.0  
 FI:T14788 - hypothetical protein DKFZP564A122.1 - human  
 C:Species: Homo sapiens (man)  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
 C:Accession: T14788  
 R:Duesterhoef, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
 Submitted to the Protein Sequence Database, August 1999  
 A:Reference number: Z18178  
 A:Accession: T14788  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-112 <DUF>  
 A:Cross-references: EMBL:AL110269  
 A:Experimental source: fetal brain; clone DKFZP564A122  
 A:Genetics:  
 A:Note: DKFZP564A122.1  
 T14788 Length: 112 February 11, 2000 15:52 Type: P Check: 6935 ..

1 MTGQYDLYLL AMDCMLQAL LMKWVFMK IDEDPYQVA PVSNGLCACAF  
 51 STDPSVLAAG THDGSVYFMA TPQVPSLOH LCRMSIRRMV PTOFVCLPFI  
 101 PSKLEFLSY RI

!!AA\_SEQUENCE 1.0  
 FI:T08722 - hypothetical protein DKFZP566D133.1 - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 13-Aug-1999  
 C:Accession: T08722  
 R:Koehrer, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
 Submitted to the Protein Sequence Database, May 1999  
 A:Reference number: Z16468  
 A:Accession: T08722  
 A:Molecule type: mRNA  
 A:Residues: 1-381 <KOP>  
 A:Cross-references: EMBL:AL050050  
 A:Experimental source: fetal kidney; clone DKFZP566D133  
 A:Genetics:  
 A:Note: DKFZP566D133.1  
 T08722 Length: 381 February 11, 2000 15:52 Type: P Check: 6844 ..

1 RMTFOFKRFR EYPTWDTIR DEEDALDEL QVIGVTSPEC LQRTGISLNI  
 51 PAPQVCISE KQSDVINAI LKQTEKEF VEKHFNDLM KAVQDEPIF  
 101 QKQSAFYIC RLILSLGNM SWDKRRSEHL LKXNEKLE LKNDSDQCR  
 151 ETHKIAVEYV AEGQEDKHSI LTNNGSOAY EDYVAGLME VNLNHCQFM  
 201 GGLQKKSIG LTPYFATST VEVIHMSR MSQSDSDSLT KKLRLGNDL  
 251 VHIWSEHTR DYRRGIIPTE PGDLVLIYIP MKNHMSIOI MKRPEVPEF  
 301 PLFDGAIYNG KVLPIVYRAT AINASRALKS LIPLYONFYE EKARLYOTIV  
 351 QHLEPTFE DFAQVSPA PYHHLPSGAD H

!!AA\_SEQUENCE 1.0  
 FI:T00702 - hypothetical protein F25965.1 - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 01-Feb-1999  
 C:Accession: T00702  
 R:Lamerdin, J.E.; McCready, P.M.; Adamson, A.W.; Burkhardt-Schultz, K.; Garcia, R.; Lamerdin, J.E.; Ramirez, M.; Stillwagen, S.; Gaines, J.; Dangnan, L.; Bruce, R.; E., Kyle, A.; Ramirez, M.; Oy, D.; Kobayashi, A.; Olsen, A.O.; Carrano, A.V.; Quan, G.; Montgomery, M.; Oy, D.; Kobayashi, A.; Olsen, A.O.; Carrano, A.V.  
 Submitted to the EMBL data library, October 1997  
 A:Description: Sequence analysis of a 1mb region in 19q13.1.  
 A:Reference number: Z14199  
 A:Accession: T00702  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-196 <LAMB>  
 A:Cross-references: EMBL:AC002398, NID:92529398, PID:92477514  
 A:Genetics:  
 A:Map position: 19  
 A:Note: Intron positions not resolved (incomplete sequence)  
 A:Note: F25965.1  
 T00702 Length: 196 February 11, 2000 15:52 Type: P Check: 4343 ..

1 SWRWSKPGIN WMLSCSYCWR RVTWTGSYWM RKLGKHPQRP TINTYVILF  
 51 DRSVDLAQFS ENTPLYPICR AMRNSPSYR ERECSRSSPL PPLPDESGS  
 101 EYVNSKSDYV YKLPPTPPG PGGDACRSRI PSLQPMQGS TPDDDESPSE  
 151 PSPSTLIYRN MQRWKRIROR WRKASHRNOL RYSESKILIR EYVENO

!!AA\_SEQUENCE 1.0  
 FI:A44478 - probable cell growth or differentiation regulator (alternatively spliced type I transcript) - human  
 C:Species: Homo sapiens (man)  
 C:Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Mar-1998  
 C:Accession: A44478  
 R:Welter, J. N.; Kasperczyk, A.; Mohan, C.; Kronitiris, T.G.  
 Genomics 14, 309-319, 1992

A>Title: The HRA51 gene cluster: two upstream regions recognizing transcripts and a third encoding a gene with a leucine zipper domain.  
 A:Reference number: A44478; MUID:93052330  
 A:Accession: A44478  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid  
 A:Residues: 1-373 <ME>  
 A:Cross-references: GB:M91083; NID:g184389; PID:g184390  
 A:Note: sequence extracted from NCBI backbone (NCBIP:117869)

A44478 Length: 373 February 11, 2000 15:52 Type: P Check: 3379

```

1 MLGGLAMEL KVVWDGICRV VCGVSEQITC QEVVIALQA IGTGRFVLV
51 QRLKEKEROL LPOECFVGAQ ATCGOFASDV QFVLRTGSPS LAGRPSSDSC
101 PPRERCLIRA SLVPRRAAL GCEPRKTLTP EPAPLSRPG PAMPYPTPG
151 CCTDLRGLEL RVQRNAEELG HEAFWEQELR REQARERGQ ARLOALSAAT
201 AEHAARLQAL DAQARALEAE LQLAAPRGP PSPMASATER LHODLAVOER
251 QSAEVOGSLA IYSRALEAAE RALQAQOEL EELNRELKQC NLOQFTQOTG
301 AALPPPRPD RGPPTGCPPL PPARRESILG APSESHAGQ PPRGGPHDA
351 ELLEVAAPA PEMCLAPQ OAL

```

11AA\_SEQUENCE 1.0  
 P1:B44478 - probable cell growth or differentiation regulator (alternatively spliced type II transcript) - human  
 C:Species: Homo sapiens (man)  
 C:Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 03-Mar-1995  
 C:Accession: B44478  
 R:Wetzel, J.N.; Kasparczyk, A.; Mohan, C.; Krontiris, T.G.  
 A:Title: The HRA51 gene cluster: two upstream regions recognizing transcripts and a third encoding a gene with a leucine zipper domain.  
 A:Reference number: A44478; MUID:93052330  
 A:Accession: B44478  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid  
 A:Residues: 1-337 <ME>  
 A:Note: sequence extracted from NCBI backbone (NCBIP:117871)

B44478 Length: 337 February 11, 2000 15:52 Type: P Check: 908

```

1 MLGGLAMEL KVVWDGICRV VCGVSEQITC QEVVIALQA IGTGRFVLV
51 QRLKEKEROL LPOECFVGAQ ATCGOFASDV QFVLRTGSPS LAGRPSSDSC
101 PPRERCLIRA SLVPRRAAL GCEPRKTLTP EPAPLSRPG PAMPYPTPG
151 CCTDLRGLEL RVQRNAEELG HEAFWEQELR REQARERGQ ARLOALSAAT
201 AEHAARLQAL DAQARALEAE LQLAAPRGP PSPMASATER LHODLAVOER
251 QSAEVOGSLA IYSRALEAAE RALQAQOEL EELNRELKQC NLOQFTQOTG
301 AALPPPRPD RGPPTGCPPL PPARRESILG APSESHAGQ PPRGGPHDA

```

11AA\_SEQUENCE 1.0  
 P1:A61188 - probable transcription factor SC1 - human  
 C:Species: Homo sapiens (man)  
 C:Date: 12-May-1994 #sequence\_revision 09-Sep-1994 #text\_change 07-May-1999  
 C:Accession: A61188  
 R:Ku, D.H.; Chang, C.; Konieczki, J.; Cammizaro, L.A.; Boghosian-Sell, L.; Alder, H.; Baserga, R.  
 A:Title: A new growth-regulated complementary DNA with the sequence of a putative trans-activating factor.  
 A:Reference number: A61188; MUID:91329275  
 A:Accession: A61188

A:Molecule type: mRNA  
 A:Residues: 1-358 <KUA>  
 A:Note: authors translated the codon TAC for residue 19 as Thr  
 C:Genetics:  
 A:Gene: SC1  
 A:Map position: 6p21-22

A61188 Length: 358 February 11, 2000 15:52 Type: P Check: 916

```

1 MLPCFOLLRI GGGRGGLTYT FHPRAGAGT YRLHRADLC DVALRPOEP
51 GLISGIAEL HAEPBGDMR VSLDEHSSQG TLVNNVLRP GHRELSDGD
101 LITFGEPP GTSPSEFYEM FQNVKYPQD FAITTPRSR GEARVAGFR
151 PMLPSGAPQ RPLSTSPAP KATILNSIG SLKLRPQPL TTPSWGSPK
201 SLVPAPCE VGTTPSAPQ RNRKSVHRV LAELDESEP PENPPVIME
251 PRKLRYDKA PLPTGNRNG RPKYPSAP MAPPAVAGT VOLVAAAPR
301 KRWGPFYVM AVTSGKMPV LAASRLPGR PTSDAAGL AFSRSTRKA
351 PSDTRAHE

```

11AA\_SEQUENCE 1.0  
 F1:S63679 - signal transducer and activator of transcription 2 type a - human  
 C:Species: Homo sapiens (man)  
 C:Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 28-May-1999  
 C:Accession: S63679  
 R:Stojanovic, T.; Nishio, Y.; Kishimoto, T.; Akira, S.  
 A:Title: Identification of alternative splicing form of Stat2.  
 A:Reference number: S63679; MUID:96176320  
 A:Accession: S63679  
 A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-65 <SVG>  
 A:Cross-references: GB:S81491; NID:g1478435; PID:NAA36225.1; PID:g1478436

S63679 Length: 65 February 11, 2000 15:52 Type: P Check: 1615

```

1 GFILRFSES SEGRTCSWV EHQDDKVL I YSVPTKEY LOSLPTEI
51 IRHYQLTEE NIPEN

```

11AA\_SEQUENCE 1.0  
 P1:A41795 - somatostatin receptor 1 - human  
 C:Species: Homo sapiens (man)  
 C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 24-Sep-1998  
 C:Accession: A41795  
 R:Yamada, Y.; Post, S.R.; Wang, K.; Tager, H.S.; Bell, G.I.; Saito, S.  
 A:Title: Cloning and functional characterization of a family of human and mouse somatostatin receptors expressed in brain, gastrointestinal tract, and kidney.  
 A:Reference number: A41795; MUID:92108031  
 A:Accession: A41795  
 A:Molecule type: DNA  
 A:Residues: 1-391 <YAM>  
 A:Cross-references: GB:M81829; NID:g307433; PID:g307434  
 A:Note: sequence extracted from NCBI backbone (NCBIP:74767, NCBIP:74768)  
 C:Genetics:  
 A:Gene: GDB:SSR1  
 A:Cross-references: GDB:134185; OMIM:182451  
 A:Map position: 14q13-14q13  
 A:Initons: #Status absent  
 A:Keywords: G protein-coupled receptor; glycoprotein; hormone receptor; lipoprotein; phosphoprotein; thiolester bond; transmembrane protein  
 F158-84/Domain: transmembrane #status predicted <TM>  
 F195-120/Domain: transmembrane #status predicted <TM>  
 F1132-153/Domain: transmembrane #status predicted <TM>

F:173-195/Domain: transmembrane #status predicted <TM4>  
 F:220-250/Domain: transmembrane #status predicted <TM5>  
 F:269-296/Domain: transmembrane #status predicted <TM6>  
 F:302-336/Domain: transmembrane #status predicted <TM7>  
 F:4.44/48.381/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:130-208/Disulfide bonds: #status predicted (Asn) (covalent) #status predicted  
 F:172/Binding site: phosphate (Thr) (covalent) (by CAMP-dependent kinase)  
 #status predicted  
 F:265/Binding site: phosphate (Ser) (covalent) (by CAMP-dependent kinase)  
 #status predicted  
 F:339/Binding site: palmitate (Cys) (covalent) #status predicted  
 A41795 Length: 391 February 11, 2000 15:52 Type: P Check: 8652 ..

1 MFPNQTASSP SSSPSPSPGS CGEGGSGRNP GAGADGME PERNASNGT  
 51 LSEGGGSAVL ISFTYSVCL VGLGNSMVI YVILRRAKM TATNIYIILN  
 101 ATADELLMLS VPFLVSTLL RHPFGALIC RLVLSDAVN METSIYCLTV  
 151 LSVDRVAVV HPKARYRR PTVAKVNLG VWVLSLIVL PIVESRTAA  
 201 NSDGIYACNM LMPEPAQRL VGFVLYTFIL GELLVGAIC LCVYLIAM  
 251 RMVALKAGW QKRSEKIT LMMVMVWVF VICWPFYVY QLVNFAEOD  
 301 DATVSOLSVI LGYANSCANP ILYGLSDNF KRSFORLICL SWMDNAEBP  
 351 VDIYATALKS RAYVEDPQP ENLESGVFR NGTCRSRTT L

11AA\_SEQUENCE 1.0  
 P1:JC5627 - STAT induced STAT inhibitor 3 - human  
 C:Species: Homo sapiens (man)  
 C>Date: 14-Oct-1997 #sequence\_revision 07-Nov-1997 #text\_change 07-May-1999  
 C:Accession: JC5627  
 R:Matsumoto, S.; Ikegami, K.; Ueno, K.; Narazaki, M.; Naka, T.; Yamamoto, H.;  
 Matsumoto, T.; Saito, H.; Hosoe, S.; Kishimoto, T.  
 Biochem. Biophys. Res. Commun. 237, 79-83, 1997  
 A:Title: Cloning and functional analysis of new members of STAT induced STAT  
 inhibitor (SSI) family: SSI-2 and SSI-3.  
 A:Reference number: JC5626; MUID:97410118  
 A:Accession: JC5627  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-225 <MIN>  
 C:Comment: This protein plays a role in negative feedback control of Janus  
 kinase-signal transduction and activator of transcription signaling pathway.  
 C:Superfamily: SH2 homology  
 F:45-144/Domain: SH2 homology <SH2>  
 F:189-197/Region: SC motif 1  
 F:210-218/Region: SC motif 2  
 JC5627 Length: 225 February 11, 2000 15:52 Type: P Check: 178 ..

1 MYTHSKFPAA GMSRPIDTSL RLKTFSSKSE YQLVNVAVR LQESGFYWSA  
 51 VTGGGANLLI SARPAGTFELI RDSPORHFF ALVKTQSGT KMLRIQCEGG  
 101 SFSLSDDPRS TQPYRPDCV LKLVHYMP PGARSPSP TPSSSEVEDO  
 151 PSAQPLPGSP PRKAYITSG GEKIPVLVSR PLSSNVALIQ HLCRTVNGH  
 201 LDSYEKVTOL PGPIREFLDQ YDAPL

11AA\_SEQUENCE 1.0  
 P1:S41063 - translocon-associated protein beta chain - human  
 C:Species: Homo sapiens (man)  
 C>Date: 25-Oct-1994 #sequence\_revision 10-Nov-1995 #text\_change 28-May-1999  
 C:Accession: S41063  
 R:Bodesscot, M.; Brison, O.  
 Biochim. Biophys. Acta 1217, 101-102, 1994  
 A:Title: Cloning and sequence analysis of the beta subunit of the human  
 translocon-associated protein.

A:Reference number: S41063; MUID:94114564  
 A:Accession: S41063  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-183 <MOD>  
 A:Cross-references: GB:X74104; NID:9452756; PIDN:CAA52207.1; PID:9452757  
 S41063 Length: 183 February 11, 2000 15:52 Type: P Check: 8093 ..

1 MRLSLFVYLA LEAVTQAEKG ARLAKSKSL NRYNVEGRDL TLQYNIYNG  
 51 SSALDVELS DDSFPPEDFG IVSGMLNVKW DRIAPASNV HTVYLRPLKA  
 101 GFNFNTSATV TYLAQEDGPV VIGSTAPQO GGTLAQRFD RRFSPHFLDW  
 151 AAFGVMTLPS IGIPILLWYS SKRKYDTPKT KKN

11AA\_SEQUENCE 1.0  
 P1:A56446 - Ig heavy chain V region (3H-3H scFv) - mouse (strain BALB/c)  
 C:Species: Mus musculus (house mouse)  
 C>Date: 19-Jan-1996 #sequence\_revision 19-Jan-1996 #text\_change 16-Aug-1996  
 C:Accession: A56446  
 R:Tang, F.M.; Poltz, L.A.; Mahoney, W.C.; Schueler, P.A.  
 J. Biol. Chem. 270, 7829-7835, 1995  
 A:Title: A high affinity dioxin-binding protein displayed on M13 is  
 functionally identical to the native protein.  
 A:Reference number: A56446; MUID:95225583  
 A:Accession: A56446  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-268 <TRAN>  
 A:Cross-references: GB:U20617

A:Reference number: A56679; MUID:9115924  
 A:Accession: A56679  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-183 <GOE>  
 A:Cross-references: GB:X53529; NID:9936; PID:9937  
 R:Wada, T.; Rindress, D.; Cameron, P.H.; Ou, W.J.; Doherty II, J.J.; Louvard,  
 D.; Bell, A.W.; Dignard, D.; Thomas, D.V.; Bergeron, J.J.M.  
 J. Biol. Chem. 266, 19599-19610, 1991  
 A:Title: SRalpha and associated calnexin are major calcium binding proteins of  
 the endoplasmic reticulum membrane  
 A:Reference number: A37273; MUID:92011761  
 A:Accession: B37273  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-183 <MAD>  
 A:Cross-references: EMBL:X53591; NID:9845; PID:9846  
 C:Keywords: endoplasmic reticulum; glycoprotein; transmembrane protein  
 A56679 Length: 183 February 11, 2000 15:52 Type: P Check: 8392 ..

1 MRLASVYLA LEAVSHAEBG ARLAKSKSL NRYNVEGRDL TLQYNIYNG  
 51 SSALDVELS DDSFPPEDFG IVSGMLNVKW DRIAPASNV HTVYLRPLKA  
 101 GFNFNTSATV TYLAQEDGPV VIGFTAPQO GGTLAQRFD RRFSPHFLDW  
 151 AAFGVMTLPS IGIPILLWYS SKRKYDTPKS KKN

Mon Feb 14 08:07:32 2000

plf.cat

Page 178

C:Keywords: heterotetramer; immunoglobulin

A56446 Length: 268 February 11, 2000 15:52 Type: P Check: 3609 ..

1 MAOVKLOESG AELVKPGASV KLSCTTSGFN IKDTYHMYK QREPGLEMI  
51 GRIAPANGIT KYDPKFOGKA TIADTSSNT AYLOLSLTS EDTAVYCAS  
101 YLTRYENYW GQGTIVTVSS GGGSGGSGS GGGSDIELT OSPALMSAL  
151 GERYTMSORA SSSVNFYIYW QOKSPASPKL WYVYTHLDP GVPARFSGSG  
201 GSNYSYLITS SMEGEDAATY YCOQFTSPF TFGSGTKLEI KRSAHHHHH  
251 HGAAEOKLIS EEDLNGAA

!!AA\_SEQUENCE 1.0  
F1:B61188 - sci protein - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 19-Mar-1997 #sequence\_revision 13-Mar-1998 #text\_change 07-May-1999  
A:Accession: B61188  
R:Ku, D.H.; Chang, C.; Koniecki, J.; Cannizzaro, L.A.; Boghosian-Sell, L.;  
Alder, H.; Baseraga, R.  
Cell Growth Differ. 2, 179-186, 1991  
A:Title: A new growth-regulated complementary DNA with the sequence of a  
putative trans-activating factor  
A:Reference number: A61188; MUID:91329275  
A:Accession: B61188  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-265 <KUA>

B61188 Length: 265 February 11, 2000 15:52 Type: P Check: 9720 ..

1 MLPQFLKMI GGGKGVLYT FQAPAKSGCT YRLCCRADLC DVALRPOQRP  
51 GLISGHAEL HAELOGDDMR VSLIEDSSOG TLVNNVLP RGRLEISDSD  
101 LITFGQGA GTSSSEPFY MFOQVVKPO DFAITVPRS KEGAGGFRP  
151 MPPGQPORP LSTLSSAPKA TLILSMGSL SKLQAOPLTF CGGGRPRGL  
201 AIPSGHGA QVSPAPQPN RRSVHKYLA EMLDEVSPE SPLSLTEPR  
251 KRLRVEKAL IASGE

!!AA\_SEQUENCE 1.0  
F1:S63681 - signal transducer and activator of transcription 2 type a - mouse  
(fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 29-Jan-1999  
R:Singh, T.; Nishio, Y.; Kishimoto, T.; Akira, S.  
FEBS Lett. 381, 191-194, 1996  
A:Title: Identification of alternative splicing form of Stat2.  
A:Reference number: S63679; MUID:96176320  
A:Accession: S63681  
A:Status: preliminary; nucleic acid sequence not shown; not compared with  
conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-64 <SUG>  
A:Cross-references: GB:S61342; NID:91478438; PID:91478439

S63681 Length: 64 February 11, 2000 15:52 Type: P Check: 840 ..

1 GAFILRSET SSGITCSMV EHODDKEI YSVQPTREV LQSLPLEII  
51 RHQVLALEN IPEN

!!AA\_SEQUENCE 1.0  
P1:C41795 - somatostatin receptor 1 - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 30-Sep-1993 #sequence\_revision 31-Dec-1993 #text\_change 10-Sep-1997

C:Accession: C41795  
R:Yamada, Y.; Post, S.R.; Wang, K.; Tager, H.S.; Bell, G.I.; Seino, S.  
Proc. Natl. Acad. Sci. U.S.A. 89, 251-255, 1992  
A:Title: Cloning and functional characterization of a family of human and mouse  
somatostatin receptors expressed in brain, gastrointestinal tract, and kidney.  
A:Reference number: A41795; MUID:92108031  
A:Accession: C41795  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-391 <YAM>  
A:Cross-references: GB:M81831; NID:9201058; PID:9201059  
C:Keywords: G protein-coupled receptor; hormone receptor; transmembrane protein

C41795 Length: 391 February 11, 2000 15:52 Type: P Check: 8110 ..

1 MFPNGTASP SSSPSPSGS CGEGACSRG GSGADGME PGNNASQNT  
51 LSEOGSAIL ISFIYSVCL VGLGNSMT YILRYAKM TATNITLM  
101 AIADELMLS VPELTSITL RHPFGALC RLVSVDVN MFTSICLV  
151 LSVDRVAVV HPIKARRR PTVAKVNIQ VWLSLVLV PIVFSRTAA  
201 NSDGTACMK LMPEAORML VGFVYTFELM GELLPGAIC LCVLILAKM  
251 RMVALAGWQ QKRSERKIT LMYMNVYVF VICMPPYV QLVNFAEOD  
301 DATVSQSLVI LGVNSCANP ILXGFLDNF KRSFORITCL SWMDNAEEP  
351 VDIYATLAKS RAYVEDFOP ENLESGVFR NGTCASRIST L

!!AA\_SEQUENCE 1.0  
P1:S66270 - kynurenine--oxoglutarate transaminase (EC 2.6.1.7) /  
asparagine--oxo-acid transaminase (EC 2.6.1.14) K precursor - rat  
N:Alternate names: kynurenine aminotransferase / glutamine transaminase K  
N:Contains: asparagine--oxo-acid transaminase (EC 2.6.1.14) K, cytosolic;  
asparagine--oxo-acid transaminase (EC 2.6.1.14) K, mitochondrial; cysteine  
conjugate beta-lyase; kynurenine--oxoglutarate transaminase (EC 2.6.1.7),  
cytosolic; kynurenine--oxoglutarate transaminase (EC 2.6.1.7), mitochondrial  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 22-Nov-1996 #sequence\_revision 22-Nov-1996 #text\_change 17-Mar-1999  
R:Mahebe, P.; Alberati-Giani, D.; Koehler, C.; Cesura, A.M.  
Submitted to the EMBL Data Library, May 1995  
A:Description: Identification of a mitochondrial form of kynurenine  
aminotransferase/ glutamine transaminase K from rat brain.  
A:Reference number: S66270  
A:Accession: S66270  
A:Molecule type: mRNA  
A:Residues: 1-457 <MAL>  
A:Cross-references: EMBL:Z49696; NID:9899479; PID:91030066  
R:Mahebe, P.; Alberati-Giani, D.; Koehler, C.; Cesura, A.M.  
FEBS Lett. 367, 141-144, 1995  
A:Title: Identification of a mitochondrial form of kynurenine  
aminotransferase/glutamine transaminase K from rat brain.  
A:Reference number: S66269; MUID:95317413  
A:Accession: S66269  
A:Molecule type: mRNA  
A:Residues: 1-71:432-457 <MAN>  
A:Cross-references: EMBL:Z49696  
R:Perry, S.J.; Schofield, M.A.; MacFarlane, M.; Lock, E.A.; King, L.J.; Gidson,  
G.G.; Goldfarb, P.S.  
Mol. Pharmacol. 43, 660-665, 1993  
A:Title: Isolation and expression of a cDNA coding for rat kidney cytosolic  
cysteine conjugate beta-lyase.  
A:Reference number: 157944; MUID:93275300  
A:Accession: 157944  
A:Molecule type: mRNA  
A:Residues: 35-457 <RES>  
A:Cross-references: GB:S61960; NID:9300515; PID:9300516  
C:Genetics:  
A:Genome: nuclear  
C:Keywords: alternative initiators; aminotransferase; mitochondrion

F.1.32/Domain: transit peptide (mitochondrion) #status experimental <TRP>  
 F.33-457/Product: kynurenine--oxoglutarate transaminase / asparagine--oxo-acid  
 transaminase K, mitochondrial #status experimental <MRL>  
 F.33-457/Product: kynurenine--oxoglutarate transaminase / asparagine--oxo-acid  
 transaminase K, cytosolic #status experimental <MA12>

566270 Length: 457 February 11, 2000 15:52 Type: P Check: 7215

1 MERSAALSY HLMPLMGRK AGASITRCLH QSLTMKRLQ ARRDGIDQN  
 51 LMVEGKLTK EYDVNLGQG PPDFSPDFA TQAFQATISG NEMNLQYRA  
 101 FGYPPLTNVL ASFFGKLIGQ EMDPLTNVL TVGAYGALF RQGLVDEGD  
 151 EYIMEPAFD CYEPMTMAG GCPVEVTLK SPAPKGLGA SMDQGLDAE  
 201 LAKFEPRTK ILVLTENNP LGRVSRMEL ELVANLCOH DVCISDEVY  
 251 QMLVYDGHQ VSIALPGMW DRLITIGSAG KSFATGKMY GMYGPDNIM  
 301 KHLRTYHNS IFHCPTQQA AVACPEREQ QHFGQSSYF LQLPQAMEIN  
 351 RDMIRSLQS VGLKWLISQG SYPLIADISD FKSRAKPLIG AEDEPYDRF  
 401 AKMINMGL VGIPSTFFS RPHQKDFDH IRFCVAKDKA TLQAMDERLR  
 451 KMKELQ

!!AA\_SEQUENCE 1.0  
 P1:S41286 - latexin - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 10-Sep-1997  
 C:Accession: S41286  
 R:Hanawa, Y.; Uraeai, Y.; Takiguchi-Hayashi, K.; Onon, A.; Satoh, K.; Miyamoto, M.; Atimatsu, Y.  
 Submitted to the EMBL Data Library, December 1993  
 A:Description: Intracellular regionally represented specific transcription for a novel protein latexin.  
 A:Reference number: S41286  
 A:Accession: S41286  
 A:Status: Preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-223 <MRL>  
 A:Cross-references: EMBL:X76985; NID:g440575; PID:g440576

S41286 Length: 223 February 11, 2000 15:52 Type: P Check: 8455

1 MEIIPPTHY ASRAAAYEN CLNQOQTPN KYKVCYVQO ASKEDIPERG  
 51 HRYHLKFSYE EIIOKVTVS CTAELVYPRM GQGSAPENVF TREGIGKNP  
 101 DEEDNTFYOR LMSKKEPLQA QNIPDNFGNV SPQKKPYNHL AMVACGYVM  
 151 ONSTEDTWK MAKIQTVKOV QRRNDFIELD YTVLHNVAS QELIPWQOV  
 201 LMHPQYGVY KHSRLPKFA PAE

!!AA\_SEQUENCE 1.0  
 P1:A39297 - somatostatin receptor - rat  
 N:Alternate names: probable G-protein-coupled receptor; SRIF receptor  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 03-Aug-1992 #sequence\_revision 03-Aug-1992 #text\_change 31-Oct-1997  
 C:Accession: A39297; A45102; S20088  
 R:Meierhof, W.; Paust, H.J.; Schoenrock, C.; Richter, D.  
 DNA Cell Biol. 10, 689-694, 1991  
 A:Title: Cloning of a cDNA encoding a novel putative G-protein-coupled receptor expressed in specific rat brain regions.  
 A:Reference number: A39297; MUID:92096119  
 A:Accession: A39297  
 A:Molecule type: mRNA  
 A:Residues: 1-391 <MEV>  
 A:Cross-references: GB:X62314; GB:X61630; NID:g56309; PID:g56310  
 A:Experimental source: brain

A:Note: It is uncertain whether Met-1 is the initiator or whether translation is initiated 5' to the sequenced region  
 R.L.I. X.J., Forte, M.; North, R.A.; Ross, C.A.; Snyder, S.H.  
 J. Biol. Chem. 267, 21307-21312, 1992

A:Title: Cloning and expression of a rat somatostatin receptor enriched in brain.  
 A:Reference number: A45102; MUID:93016064  
 A:Accession: A45102  
 A:Status: Preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid  
 A:Residues: 1-391 <DLI>  
 A:Experimental source: brain  
 A:Note: sequence extracted from NCBI backbone (NCBI:116692)  
 C:Keywords: G protein-coupled receptor; glycoprotein; receptor; transmembrane protein

A39297 Length: 391 February 11, 2000 15:52 Type: P Check: 9240

1 MFPNGTASP TSPPSSPPG CGEGYCSRG GSGADGMEP PGRNSQNGT  
 51 LSEGGASALT ISFTYSVCL VGLGNSMYI YVILRAKKK TATNIYIINL  
 101 AINDELMLS VPELVISTL RHPFGALLC RLVLSVDAV MFTSYCLTV  
 151 LSYDRYVAVY HPKARFRR PTVAKVNLG VWVLSLVIL PIVFSRTAA  
 201 NSDGTACNM LMPPAQRWL VGFVLTFLM GFLIPVGAIC LCVYLIAKM  
 251 RMVALRAGNQ QKRSEKIT LVMVMVYVF VICMPPYV QLVNFAEQD  
 301 DATVSOLSVI LGVANSCLNP ILVGLSDNF KRSPRILCL SMDNNAEPP  
 351 VDYATATLKS RAYVEDPQ ENLESGVFR NGTCASRIST L

!!AA\_SEQUENCE 1.0  
 P1:JC4294 - ammonia monoxygenase (EC 1.-.-.-) A chain - Nitrosospira sp.  
 C:Species: Nitrosospira sp.  
 C>Date: 14-Nov-1995 #sequence\_revision 08-Feb-1996 #text\_change 21-Mar-1996  
 C:Accession: JC4294  
 R:Klotz, M.G.; Norton, J.M.  
 Gene 163, 159-160, 1995  
 A:Title: Sequence of an ammonia monoxygenase subunit A-encoding gene from Nitrosospira sp. NPAV.  
 A:Reference number: JC4294; MUID:96001263  
 A:Accession: JC4294  
 A:Molecule type: DNA  
 A:Residues: 1-274 <KLO>  
 A:Cross-references: GB:U20644  
 A:Experimental source: NPAV  
 C:Keywords: monooxygenase; oxidoreductase  
 A:gene: amoa

JC4294 Length: 274 February 11, 2000 15:52 Type: P Check: 404

1 MSRTDEILKA AKMPPEVYM SMIDAIYFP ILCLILVGY HMFMLLAGD  
 51 WDFWLMKDR QMPPVTPPIV GITTCATIMY DLWVNRLPF GATLICVCLL  
 101 VGEMLTRFNG FYWMSHPIN FVLPSMTIPG ALIMDTYML TRNMNITLV  
 151 GGGARGLLY PGNMPPGPT HPLVAGEVL LSLADYGLF YVRTGTPEV  
 201 RLIEQSLRT FGHRTIYAG FFSAFVSLM FCVMWYFGKL YCTAFYVKG  
 251 PGRVTRKND VTAGEBCFP EGIR

!!AA\_SEQUENCE 1.0  
 P1:C75042 - aspartate carbamoyltransferase, catalytic chain (pyrB) PAB1498 -  
 Pyrococcus abyssi (strain Otsay)  
 C:Species: Pyrococcus abyssi  
 C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
 C:Accession: C75042

```

R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal
chromosome structure and evolution.
A:Reference number: A7501
A:Accession: C75042
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-308 <KAM>
A:Cross-references: GB:AJ248287; GB:AL096836; NID:g5458657; PIDN:CA550232.1;
PID:g516130; PID:g5458745
A:Experimental source: strain Orsay
A:Genetics:
A:Gene: PAB1498

C75042 Length: 308 February 11, 2000 15:52 Type: P Check: 9304 ..

1 MDWGRGVIS IDFSKEDIE TVLATAERLE RELKEGQLE YAKKIATL
51 FFEESTIRTL SFESAMHRLG GAVIGFAEAS TSSVKKGESL RDTIKTYEQY
101 CDVIVIRHPK EGAARLAAEV AEVPIVAGD GSNQHPOTL LDLYTIKEF
151 GRIDGLKIGL LGDLKYGRTV HSLAEALTFY DVELILISPE LLMRPHRIVE
201 ELREKGMKV ETTLEDVIG KLDVLYTRI QKERPPDEQ YLKVKSQYQ
251 NLKYLEKARD ELRIHNPDR VDEIHPEVDN TKHAIYFQV ENGVPVPMAL
301 LALVLGYI

!!AA_SEQUENCE 1.0
P1:D71333 - conserved hypothetical protein TP0373 - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 17-Mar-1999
A:Accession: D71333
R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson,
R.; Gwin, M.; Hickey, E.K.; Clayton, R.; Ketchum, K.A.; Sodergren, E.;
Richardson, D.; Howell, J.K.; Salzberg, S.; Peterson, J.; Khaliq, H.;
Artach, P.; Bowman, C.; Cotton, M.D.; Fujii, C.; Garland, S.; Hatch, B.;
Horst, K.; Roberts, K.; Watthey, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis
spirochete.
A:Reference number: A71250; MUID:98332770
A:Accession: D71333
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-477 <COL>
A:Cross-references: GB:AE001216; GB:AE000520; NID:g3322647; PID:g3322651
A:Experimental source: strain Nichols
A:Genetics:
A:Gene: TP0373

D71333 Length: 477 February 11, 2000 15:52 Type: P Check: 4876 ..

1 MSESROKILP LTVHARSFG HFLVRRPSC LTVANSGGD SLALTYAHE
51 LAPDFGVAC AVTVDSIRA QEGALDARFV RALCARFSP LPFCVQOISA
101 GAVHACAKIR GRGVODARA LRYKFDHVA ARCGAVVLT AHTRDOYER
151 LLMRLFOGAA ASALOGIRAA RGRVRLPLK VSRVCEPFL QTVGNRRD
201 ASNTCKRYR NRIRHPIPA LDAVLGWS GLDKTFAGIS AHSFCAVAL
251 TWRREGCSHA WEPVRLGT RLMRPSDFL AAEFLRFL LQACVRLGV
301 SHRVPGALE RCARFDVRR IHVSGQLR AGATVLSICI HASDARETK
351 KODAGSPSS EKQGVSAIV ARGATPCAC GTLIVEVRA GVFCVCAODH

401 VGVGPFSPF YIRHRTGDT IIRNGHKGI RKMSEWHVP LSDRTVPLMI
451 EDDGVLRALY GALLGYONRY AERTPHE

!!AA_SEQUENCE 1.0
P1:G71287 - conserved hypothetical protein TP0730 - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 17-Mar-1999
A:Accession: G71287
R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson,
R.; Gwin, M.; Hickey, E.K.; Clayton, R.; Ketchum, K.A.; Sodergren, E.;
Richardson, D.; Howell, J.K.; Salzberg, S.; Peterson, J.; Khaliq, H.;
Artach, P.; Bowman, C.; Cotton, M.D.; Fujii, C.; Garland, S.; Hatch, B.;
Horst, K.; Roberts, K.; Watthey, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis
spirochete.
A:Reference number: A71250; MUID:98332770
A:Accession: G71287
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-365 <COL>
A:Cross-references: GB:AE001245; GB:AE000520; NID:g3323033; PID:g3323035
A:Experimental source: strain Nichols
A:Genetics:
A:Gene: TP0730

G71287 Length: 365 February 11, 2000 15:52 Type: P Check: 2410 ..

1 MARASSPLSS MSHPSLPFS AHCAARLRK TCQSGSDASF OYKAREKST
51 LPAPARRRF PVQYFPSPD KRWILGYFSP PSMKERTAGI LCSYTLFMSA
101 LVLAISLHAG RTQVPPSSTL SEFLVILMH CLIEFLVAV SNEPDTIVP
151 FXPDEQAPR ICAAAASDCK ENRTALKTLL TATHILIRA SAIPVIGTL
201 KHALAGLSY ELVAGLSVLF LDFIDGRIA RARETSRNG ETLDAASYA
251 LIGLISALY QSGVPLMEF VLIITRSLQ TVIACYALF GHPMTGSTAG
301 GKATVAVTML LYTLELRL LPNLARSNSG ARRTGAEIL AGVITFTGIV
351 EKLYGVQHR PGRSP

!!AA_SEQUENCE 1.0
P1:J00857 - hypothetical 18.1K protein - Escherichia coli retron Ec67
N:Alternate names: regulatory protein CII homolog
C:Species: Escherichia coli retron Ec67
C>Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 10-Sep-1997
A:Accession: J00857
R:Hsu, M.Y.; Inouye, M.; Inouye, S.
Proc. Natl. Acad. Sci. U.S.A. 87, 9454-9458, 1990
A:Title: Retron for the 67-base multicopy single-stranded DNA from Escherichia
coli: a potential transposable element encoding both reverse transcriptase and
dam methylase functions.
A:Reference number: J00851; MUID:91067724
A:Accession: J00857
A:Molecule type: DNA
A:Residues: 1-169 <HSD>
A:Cross-references: GB:M55249; NID:g145143; PID:g145147
A:Experimental source: E. coli strain Cl-1
A:Genetics:
A:Note: Insertion site is equivalent to 19 min of E. coli K12 genetic map

J00857 Length: 169 February 11, 2000 15:52 Type: P Check: 6140 ..

1 MPDYOVSKHP HFDEACRAFA LRHNLVOLAE RAGNVOIIR NKUNSOYHL
51 LTAPEIWLIT DLEEDSTLYD GFIAQHCLP CVPINVAKK KLPHVMSAT
101 AEIGRYAAGA VSGDVYTSAG RDAISSINS VTRIMALAV SLQARLQANP

```



```

11AA_SEQUENCE 1.0
P1:T10288 - hypothetical protein 19 - Orygia pseudotsugata nuclear polyhedrosis
virus
C:Species: Orygia pseudotsugata nuclear polyhedrosis virus, OPMNPV
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C:Accession: T10288
R:Ahrens, C.A.; Russell, R.R.; Funk, C.J.; Evans, J.; Harwood, S.; Rohmann,
G.F.
Virology 229, 381-399, 1997
A:Title: The sequence of the Orygia pseudotsugata multinucléocapsid nuclear
polyhedrosis virus genome.
A:Reference number: 217011
A:Accession: T10288
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-298 <AHR>
A:Cross-references: EMBL:U75930; NID:92934903; PID:g1911265
T10288 Length: 298 February 11, 2000 15:52 Type: P Check: 562

1 MLVQVNFLO IVLHNAFLGL CSFAVFPALM ATVATARAFL LEEDSABSI
51 INLSHLAFL LGPYLATIT WAMTKMLCY KGLEKRSNFY MKTVLALHL
101 MAGSCMLLFV VFQPIHNG HVPYLDLIR HHDROSLGMS GYVVOEEVH
151 DANIRFDLN CYYDNFMKK CVGCRMEVRH DEPTVFONO GALTMLALLA
201 IYMCWNNYV QOKETRRKP RNNTITL METEKYDIA EEEHESNMR
251 SMLSRARAR NRPHNFFPC RQSGOTONS RLIFQKRRR RQVPTAE

11AA_SEQUENCE 1.0
P1:G72757 - hypothetical protein APE0052 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: G72757
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.;
Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.;
Funahashi, T.; Tanaka, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
Aoki, K.; Kubota, K.; 83-101, 1999
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1.
A:Reference number: A72450; MUID:99310339
A:Accession: G72757
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-175 <KAW>
A:Cross-references: DDBJ:AP000058; NID:g5103388; PIDN:BAA78961.1; PID:d1042737;
PID:g5103440
A:Experimental source: strain K1
A:Gene: APE0052
G72757 Length: 175 February 11, 2000 15:52 Type: P Check: 9964

1 MSPALIRTA IAAVLAVFI AICAVAIYTS EYSDVSTLOS VTRASRTVQ
51 AGVAYIGCT ATVIYGKTY TLEAGAYCI LMPIDGSGS VAFVMECEK
101 GYVAALYEL DSFTARAGS PYEDIVVVD GYVRGEELV LITPAGEESL
151 PVTVAAILK GCHAYDSEK AVEVO

11AA_SEQUENCE 1.0
P1:C72759 - hypothetical protein APE0064 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: C72759
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.;
Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.;
Funahashi, T.; Tanaka, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
Aoki, K.; Kubota, K.; 83-101, 1999
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1.
A:Reference number: A72450; MUID:99310339
A:Accession: C72759
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-114 <KAW>
A:Cross-references: DDBJ:AP000058; NID:g5103388; PIDN:BAA78973.1; PID:d1042749;
PID:g5103452
A:Experimental source: strain K1
A:Gene: APE0064
C72759 Length: 114 February 11, 2000 15:52 Type: P Check: 3453

1 MTSLSACSLA SPLWYTRST APEAKRPA ASSIRAASSG VRCVGSRS
51 TALEPRARAL ITSSLISEN PISFKYLSGS INPRASDAS ASIYACLRP
101 AFLPATNRYT ARSS

11AA_SEQUENCE 1.0
P1:A72773 - hypothetical protein APE0172 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: A72773
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.;
Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.;
Funahashi, T.; Tanaka, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
Aoki, K.; Kubota, K.; 83-101, 1999
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1.
A:Reference number: A72450; MUID:99310339
A:Accession: A72773
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-129 <KAW>
A:Cross-references: DDBJ:AP000058; NID:g5103388; PIDN:BAA79083.1; PID:d1042859;
PID:g5103562
A:Experimental source: strain K1
A:Gene: APE0172
A72773 Length: 129 February 11, 2000 15:52 Type: P Check: 7998

1 MSPLISIR GADARVSTS LFSTGFANR VTCICMNV GTATRAVMS
51 SGGSTVLTST AWMVGRVGL RTIESIGPEL TSLGAPLIT RLTLITSPN
101 SLFRVHSIR RPYVLDHTR RLPLILSS

11AA_SEQUENCE 1.0
P1:C72775 - hypothetical protein APE0189 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: C72775
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.;
Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.;
Funahashi, T.; Tanaka, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
Aoki, K.; Kubota, K.; 83-101, 1999
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1.

```

Mon Feb 14 08:07:32 2000

pir.cat

Page 182

A:Reference number: A72450; MUID:99310339  
A:Accession: C72775  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-100 <KAW>  
A:Cross-references: DDBJ:AP000058; NID:95103388; PIDN:BA79101.1; PID:d1042877;  
A:Experimental source: strain K1  
A:Gene: APE0189  
C72775 Length: 100 February 11, 2000 15:52 Type: P Check: 1302 ..  
1 MLEFRSTSD IIFDLKLAP VIIIPMSIS VSRSPILCS TLVIMKAM  
51 ASSSVASCFR DILACAVSR MRASCLVYM GTGAPSMLE WRRALLIST  
11AA\_SEQUENCE 1.0  
P1:C72784 - hypothetical protein APE0258 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C:Accession: C72784  
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, F.; Takahashi, M.; Sekine, M.; Baba, S.; Anka, A.; Kosugi, H.; Hosoyama, A.; Funahashi, T.; Tanaka, T.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.; Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix K1.  
A:Reference number: A72450; MUID:99310339  
A:Accession: C72784  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-313 <KAW>  
A:Cross-references: DDBJ:AP000058; NID:95103388; PIDN:BA79173.1; PID:d1042949;  
A:Experimental source: strain K1  
A:Gene: APE0258  
C72784 Length: 313 February 11, 2000 15:52 Type: P Check: 9325 ..  
1 MWLVGSLDD VKEVSRLG DKDFQGLMK SGLNOKKA ELLSHPEIL  
51 EAARKTEYK RESYKRLGEL VERAVESLR VGANPYIAET AEDAREIYK  
101 IYVSSSIYVM SKSWAEIG LREHLESMGN EWEETDLOGL LVMEGYKPM  
151 HNVAPAVHMT RERARIRRE KLGIEVDERR VEGNVAAYRE FLREKFFKAR  
201 VGITGANSMS ADTGITIVLE NEGNIRLVSS LPPVVAALVP IDKIVPSVMD  
251 AVNVALVOAA FAGFMFVYI SLITGPSATG DIEQIKVLGA HGREVAVVL  
301 LDNGMKRAAS HPE  
11AA\_SEQUENCE 1.0  
P1:H72721 - hypothetical protein APE0313 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C:Accession: H72721  
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, F.; Takahashi, M.; Sekine, M.; Baba, S.; Anka, A.; Kosugi, H.; Hosoyama, A.; Funahashi, T.; Tanaka, T.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.; Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix K1.  
A:Reference number: A72450; MUID:99310339  
A:Accession: H72721  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-168 <KAW>  
A:Cross-references: DDBJ:AP000059; NID:95103911; PIDN:BA79387.1; PID:d1043173;  
A:Experimental source: strain K1

A:Molecule type: DNA  
A:Residues: 1-168 <KAW>  
A:Cross-references: DDBJ:AP000059; NID:95103911; PIDN:BA79387.1; PID:d1043054;  
A:Experimental source: strain K1  
A:Gene: APE0313  
H72721 Length: 168 February 11, 2000 15:52 Type: P Check: 4962 ..  
1 MWSGEIVLR PIGFVSHGLP DSTVRENIGG VEGFIEVPE YARGDGLAG  
51 FSHIIVAYL HSKGRALRV RPRGLRLGF REEDLEPGCV FATDPSRPN  
101 PLALTIARLL DIEGRRLRVS GLDLYDQTPY LDIKYTPSR RVEDLSLPSW  
151 YKELLEASR RDPVRREV  
11AA\_SEQUENCE 1.0  
P1:H72731 - hypothetical protein APE0393 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C:Accession: H72731  
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, F.; Takahashi, M.; Sekine, M.; Baba, S.; Anka, A.; Kosugi, H.; Hosoyama, A.; Funahashi, T.; Tanaka, T.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.; Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix K1.  
A:Reference number: A72450; MUID:99310339  
A:Accession: H72731  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-172 <KAW>  
A:Cross-references: DDBJ:AP000059; NID:95103911; PIDN:BA79348.1; PID:d1043134;  
A:Experimental source: strain K1  
A:Gene: APE0393  
H72731 Length: 172 February 11, 2000 15:52 Type: P Check: 2984 ..  
1 MGSILSSSI PISSSVYLA I SSAALAIIPS VMARDLTSRA PLKTPGNTRA  
51 LTLWGRSL PVATTLAPAS LASHGHISGV GFASANTMGS LDMPLHSFL  
101 TTPGLEANA TRASAPLRAS SKLLGASPP PSVSRASLYL YLLEKRSLSI  
151 SSRSGWRPL LSTTMTLGL TP  
11AA\_SEQUENCE 1.0  
P1:G72736 - hypothetical protein APE0429 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C:Accession: G72736  
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, F.; Takahashi, M.; Sekine, M.; Baba, S.; Anka, A.; Kosugi, H.; Hosoyama, A.; Funahashi, T.; Tanaka, T.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.; Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix K1.  
A:Reference number: A72450; MUID:99310339  
A:Accession: G72736  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-160 <KAW>  
A:Cross-references: DDBJ:AP000059; NID:95103911; PIDN:BA79387.1; PID:d1043173;  
A:Experimental source: strain K1

Mon Feb 14 08:07:32 2000

pir.cat

Page 183

```
C:Genetics:
A:Gene: APE0429
G72736 Length: 180 February 11, 2000 15:52 Type: P Check: 2336 ..

1 MSASVARRRR ESKKEGEGE KSKDKLIISV EYTRDKVKV NFRONALYTL
51 IDRLGPIHER TLQITAEHQ QRGALGYER KIVAGVPSF EFKNDLIALA
101 YVGVEVNP RRNRLOTTND GKALEKHGA PKGVVELEK HYEIHNIS
151 LEDKRVDEHL KTIKOLKGR RREFGGLGF

!!AA_SEQUENCE 1.0
PI:D72737 - hypothetical protein APE0434 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: D72737
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Anka, A.; Kosugi, H.; Hosoyama, A.;
Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1.
A:Reference number: A72450; MUID:99310339
A:Accession: D72737
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-120 <RAW>
A:Cross-references: DDBJ:AP000059; NID:95103911; PIDN:BAW992.1; PID:dl043178;
A:Experimental source: strain K1
A:Genetics:
A:Gene: APE0434
D72737 Length: 120 February 11, 2000 15:52 Type: P Check: 7680 ..

1 MRPSFWSY LYFINLSE NLGCGGTATG PPTGASLR SYLPHLHF
51 LRSHONLSP QPPLSLRR ILRPVWRSS WLLQSSPSP SPASCSCSPA
101 APLCTWGRRR CISMILLSL

!!AA_SEQUENCE 1.0
PI:A72741 - hypothetical protein APE0458 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: A72741
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Anka, A.; Kosugi, H.; Hosoyama, A.;
Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1.
A:Reference number: A72450; MUID:99310339
A:Accession: A72741
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-160 <RAW>
A:Cross-references: DDBJ:AP000059; NID:95103911; PIDN:BAW9421.1; PID:dl043207;
A:Experimental source: strain K1
A:Genetics:
A:Gene: APE0458
A72741 Length: 160 February 11, 2000 15:52 Type: P Check: 338 ..

1 MWSGAILTA ETPAPITRR LKLDSSQA RASIMGSIYF SLSITSPIT

!!AA_SEQUENCE 1.0
PI:G72640 - hypothetical protein APE0559 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: G72640
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Anka, A.; Kosugi, H.; Hosoyama, A.;
Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1.
A:Reference number: A72450; MUID:99310339
A:Accession: G72640
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-147 <RAW>
A:Cross-references: DDBJ:AP000060; NID:95104188; PIDN:BAW9527.1; PID:dl043313;
A:Experimental source: strain K1
A:Genetics:
A:Gene: APE0559
G72640 Length: 147 February 11, 2000 15:52 Type: P Check: 9276 ..

1 MWTTPYPRAA STASTRPGS HGASMTPTPL CLAMTSISLT ALAASPSSSS
51 WSISGTPEP PILTPLETF HSTAAAPAR ACLQASTEPV LDSEKTPFIT
101 SSSGSPPEP AIRRAAPST SSAAALATRA RVLETGSLAG ADKLLHP

!!AA_SEQUENCE 1.0
PI:A72651 - hypothetical protein APE0638 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: A72651
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Anka, A.; Kosugi, H.; Hosoyama, A.;
Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1.
A:Reference number: A72450; MUID:99310339
A:Accession: A72651
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-119 <RAW>
A:Cross-references: DDBJ:AP000060; NID:95104188; PIDN:BAW9609.1; PID:dl043395;
A:Experimental source: strain K1
A:Genetics:
A:Gene: APE0638
A72651 Length: 119 February 11, 2000 15:52 Type: P Check: 639 ..

1 MMTRLDPPA GAGGIYSRC LSAITPSTLA GFRASARTSK ALMGFWPQ
51 LYTSTVANYL SGHWYTWML SRIMATTTLP SPIQDLSR TVMPSRAAS
101 RTTSSSSGL FRSSVLAHS

!!AA_SEQUENCE 1.0
PI:D72672 - hypothetical protein APE0802 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
```

```
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C/Accession: D72672
R:Kawarabayashi, Y.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.;
Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushiida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Seko, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A>Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1
A:Reference number: A72450; MUID:99310339
A:Accession: D72672
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-129 <KAW>
A:Cross-references: DDBJ:AP000060; NID:95104188; PIDN:BA479780.1; PID:d1043566;
PID:95104465
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE0802

D72672 Length: 129 February 11, 2000 15:52 Type: P Check: 4161 ..

1 MYPIEKSLPS QLSCLNSL EMSNFTSRSE PSSPISPENL GTLLHGLVTL
51 NPFLGLGMLT TINFALLRV STDVSTARE MLRASGROSV SSSIQAELIE
101 TDLNPLRAAS RAPKVPLEK KSYIPRLGP

11AA SEQUENCE 1.0
P1:E72679 - hypothetical protein APE0857 - Aeropyrum pernix (strain K1)
C/Species: Aeropyrum pernix
C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C/Accession: E72679
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.;
Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushiida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Seko, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A>Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1
A:Reference number: A72450; MUID:99310339
A:Accession: E72679
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-234 <KAW>
A:Cross-references: DDBJ:AP000060; NID:95104188; PIDN:BA479837.1; PID:d1043623;
PID:95104522
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE0857

E72679 Length: 234 February 11, 2000 15:52 Type: P Check: 8681 ..

1 MLINRKNQ VEEGCPROGG QDELEGVYH PPTVAPORL AGAEMDLYE
51 PVDAGAEIV EMHAGYPLVL PVPAPRQAR ILRYKAVHGL ARRRVYHRS
101 QPLDVELTLD VGAVDLGVA GGGRMHLEE RRODCSEDE SRSYLITPP
151 PKRLCVNRRR PEERKNGEG VEAVDSLCPA SRGCPLSIGA VDRLPORAAG
201 RVTHRRPPR GSRVALLPHY NGPGVAPVIO RGGG

11AA SEQUENCE 1.0
P1:F72685 - hypothetical protein APE0902 - Aeropyrum pernix (strain K1)
C/Species: Aeropyrum pernix
C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C/Accession: F72685
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.;
Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushiida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Seko, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A>Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1
A:Reference number: A72450; MUID:99310339
A:Accession: F72685
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-228 <KAW>
A:Cross-references: DDBJ:AP000060; NID:95104188; PIDN:BA479886.1; PID:d1043672;
PID:95104571
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE0902

F72685 Length: 228 February 11, 2000 15:52 Type: P Check: 9005 ..

1 MEALYRSVEQ LVLNMMGDY LGLSDEIRD VLEPIVLEAA SKLSRPAEC
51 IARKIIQKH LFLKALASRL AESYDRITRE RLEPIVONAP EIGKAPAL
101 YDVARRLGAD DIVEELRLM EAVGSPTKAR CPKCGFKALT PELMCLVCKR
151 TSSEEFKEK IGFQGLERW ASRAPEIVE EYLRSSIIY EDGALAAPSE
201 PKTPLAIVPLT LGSREKSIYK KRLERLR

11AA SEQUENCE 1.0
P1:H72607 - hypothetical protein APE1326 - Aeropyrum pernix (strain K1)
C/Species: Aeropyrum pernix
C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C/Accession: H72607
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.;
Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushiida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Seko, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A>Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1
A:Reference number: A72450; MUID:99310339
A:Accession: H72607
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-387 <KAW>
A:Cross-references: DDBJ:AP000061; NID:95104821; PIDN:BA480318.1; PID:d1044104;
PID:95105004
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1326

H72607 Length: 387 February 11, 2000 15:52 Type: P Check: 1372 ..

1 MAYSMDSYKS LVDAVHIKRL VAEELRRRVG EEGFRASRD WHARRRPIPC
51 GMTHIGVGC SYGCLYCIY DMGFTGKPOD YPLSDEIAL ALALNPYAP
101 GPKGTLIAFG SYTEPFMRKT AMRAIEYLS TRWIGNPOQ ISTKIALRG
151 LLERFIEAD PRIDYLVMT TLSRWALBP GASPEERIE FMGRLEAGL
201 SATLFRLPII PGVTDREAED ILSRAARAG NNVVIGTLRV AEGILRLRA
251 SGAVEMGEIE RLRLPWRKG EQLPIYSRD KERISKARE MGRKVLPAAC
301 SANVESHGCG CAACRLSPCG DLSKPSAGE REVALLLEAL GLPVPKYSV
351 GHSVEAVLKG GRREAELASY WIIGLLRRYP RVRTIVN

11AA SEQUENCE 1.0
P1:E72626 - hypothetical protein APE1469 - Aeropyrum pernix (strain K1)
C/Species: Aeropyrum pernix
```

```

C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: E72626
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.;
Funaki, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1.
A:Reference number: A72450; MUID:99310339
A:Accession: E72626
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-146 <KAM>
A:Cross-references: DDBJ:AP000061; NID:95104821; PIDN:BA80467.1; PID:d1044253;
PID:95105153
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1469

E72626 Length: 146 February 11, 2000 15:52 Type: P Check: 6841 ..

1 MSLVMDRFE ISLVPRFTM RGTPAYLRL ATRALAVALD RLVTGMSRL
51 MLEQMRQMW RELQSEASC NAEPSGLNL MFTAKALPG VMLAAVPALS
101 LTVVALLSRS RMLARALISM SAAPLSGLEE WPSMPSHSIS IYRLPL

11AA_SEQUENCE 1.0
P1:D72630 - hypothetical protein APE1499 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: D72630
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.;
Funaki, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1.
A:Reference number: A72450; MUID:99310339
A:Accession: D72630
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-152 <KAM>
A:Cross-references: DDBJ:AP000061; NID:95104821; PIDN:BA80499.1; PID:d1044284;
PID:95105184
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1499

D72630 Length: 152 February 11, 2000 15:52 Type: P Check: 4651 ..

1 MSLRSRSRKR RSMRGTSST IYTLAIPGLI LGWSLSCIMS TLHLFSAIT
51 SRTLSWPMG SCITPVILAI LPSLARPTSI TFSMLVSASF PLITITRVLP
101 LGTTIARRA ATPAPAPSA RIIDLRRRY MALAISCSST ATTLSTRCT
151 MG

11AA_SEQUENCE 1.0
P1:H72538 - hypothetical protein APE1597 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: H72538
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.;
Funaki, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1.
A:Reference number: A72450; MUID:99310339
A:Accession: E72537
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-175 <KAM>
A:Cross-references: DDBJ:AP000062; NID:95105244; PIDN:BA80597.1; PID:d1044383;
PID:95105284
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1597

H72538 Length: 175 February 11, 2000 15:52 Type: P Check: 9362 ..

1 MDECHRLVGY IYIYLANMY RVLGSRLLHA HRHRDRKV NPVTPACRTV
51 VMLGAGTGD GEVEYLLIGY VASVYVNPV VVGRRYGGQI QYQALRLLD
101 RKVVAHNSNL AADDRHGGE CYGLPVPHPH NLHLDPGRR PLGLPAVED
151 PLSLENKPLQ GGERGVAGV AGVVG

11AA_SEQUENCE 1.0
P1:D72545 - hypothetical protein APE1648 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: D72545
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.;
Funaki, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1.
A:Reference number: A72450; MUID:99310339
A:Accession: D72545
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-194 <KAM>
A:Cross-references: DDBJ:AP000062; NID:95105244; PIDN:BA80649.1; PID:d1044435;
PID:95105336
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1648

D72545 Length: 194 February 11, 2000 15:52 Type: P Check: 2785 ..

1 MWGRGLKWT LGDRVLAVSM GDLTKVRAEA VVNPANSLMI MGGAGAGALK
51 RAGGSVIEE AMRKAPYVG EAVITSGSL PARFVHAFT MEEPMRIFPL
101 VNAFRASYAA LRLSEAGIE SVAMPAGAG VGLSLVAEVA REAMAAASIL
151 RGMWPRTIIL VARGEAYRG MEKGVREALG VEGGECPADL ARLV

11AA_SEQUENCE 1.0
P1:E72557 - hypothetical protein APE1745 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: E72557
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.;
Funaki, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1.
A:Reference number: A72450; MUID:99310339
A:Accession: E72557

```

Mon Feb 14 08:07:32 2000

pir.cat

Page 186

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1132 <KAW>  
A:Cross-references: DDBJ:AP000062; NID:95105244; PIDN:BAA80746.1; PID:d1044532;  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE1745

E72557 Length: 132 February 11, 2000 15:52 Type: P Check: 8347

1 MRVLLPGRFS LTASSRYMAL EPESILITSA CRINERAAVY TFFSNAPRP  
51 TLPGTMTISP SLACFLILR LLSALLLNGL SRSATLSQT GAEFLAALL  
101 SSPISLRVTR LVGPVGLMTP LHPSLRLQ TS

!!AA-SEQUENCE 1.0  
P1:H72557 - hypothetical protein APE1748 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.; Funahashi, T.; Tanaka, T.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.; Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.  
DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix K1  
A:Reference number: A72450; MUID:99310339  
A:Accession: H72557  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1181 <KAW>  
A:Cross-references: DDBJ:AP000062; NID:95105244; PIDN:BAA80749.1; PID:d1044535;  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE1748

H72557 Length: 181 February 11, 2000 15:52 Type: P Check: 8142

1 MFFSSVLLG SPTSMVLSGS KRAVRSLLI LSIILSSSS LSHRNKSTR  
51 PRAIILIAS AMPITPPPT LTMSTRLLT PSPAIRGSI IFILAIISGY  
101 ISRGPILOT LPLGLITAL AIAFLPVE ITILSTAPC PLTNSQHPSS  
151 TRAAQRHQA LCSESIAMRH RPRGSTRPC L

!!AA-SEQUENCE 1.0  
P1:D7257 - hypothetical protein APE1900 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.; Funahashi, T.; Tanaka, T.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.; Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.  
DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix K1  
A:Reference number: A72450; MUID:99310339  
A:Accession: D72577  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1100 <KAW>  
A:Cross-references: DDBJ:AP000062; NID:95105244; PIDN:BAA80905.1; PID:d1044691;  
A:Experimental source: strain K1  
C:Genetics:

A:Gene: APE1900

D72577 Length: 100 February 11, 2000 15:52 Type: P Check: 2318

1 MIPVVFINT SMSEFGKSO TAMEPVLRSS ILSGHNISM DLTTCSSIP  
51 SLSFNNFS LKLSATMSI EIKWPIISA RTATRPPIR APDIVETPI

!!AA-SEQUENCE 1.0  
P1:A72511 - hypothetical protein APE2063 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.; Funahashi, T.; Tanaka, T.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.; Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.  
DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix K1  
A:Reference number: A72450; MUID:99310339  
A:Accession: A72511  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1140 <KAW>  
A:Cross-references: DDBJ:AP000063; NID:95105654; PIDN:BAA81073.1; PID:d1044859;  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE2063

A72511 Length: 140 February 11, 2000 15:52 Type: P Check: 2835

1 MLSASSMS DMKRTIATA SRASLTSLT LPFLSNLSPK AFSNTSAGP  
51 LNSMLWYLF SHSPSLIATI PALTRAGWS SASSGCLPFS LHSLLDSNI  
101 LHSKAGMPS LAPISQLRFQ LNTSRPMSI VYNATASSSS

!!AA-SEQUENCE 1.0  
P1:C72529 - hypothetical protein APE2207 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.; Funahashi, T.; Tanaka, T.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.; Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.  
DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix K1  
A:Reference number: A72450; MUID:99310339  
A:Accession: C72529  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1310 <KAW>  
A:Cross-references: DDBJ:AP000063; NID:95105654; PIDN:BAA81219.1; PID:d1045005;  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE2207

C72529 Length: 310 February 11, 2000 15:52 Type: P Check: 9995

1 MRAEVRVTF SDIVDIYAYN YFSWLAVALV KPELEBRNKL IAGIQVYPM  
51 YNSRLILST AVLLGYFS FWTLGFSI IYKAIYVLLA LSAFPMFAA  
101 GLMPSLKTG SRKEGIETEL PFFAAYLSIM GRGQVPSVLY IDRAVSLKIF  
151 KALREAEEMI RTKIRILGKN PLDALEHVL DSSSVFNF ILGTTAVKV

Mon Feb 14 08:07:32 2000

pir.cat

Page 187

201 GSDVHYLEI RTODLFERRL NDIRLAEEN TLLEYIYV AVIATIVERI  
251 FTISVFEHG NLGVASAOY ASTTOLILIS FLVLEPIYVI MLVLDYRBP  
301 KEPIQIKDTW

11AA\_SEQUENCE 1.0  
P1:F72471 - hypothetical protein APE2415 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C:Accession: F72471  
R:Accession: Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba, S.; Anket, A.; Kotsugi, H.; Hosoyama, A.; Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, J.; Kusuda, M.; Oguchi, A.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, Y.; Nomura, N.; Sako, Y.; Kikuchi, H. Aoki, K. 6, 83-101, 1999  
A:Title: The complete genome sequence of an aerobic hyper-thermophilic DNA Res. Complete genome sequence of an aerobic hyper-thermophilic A:Title: Aeropyrum pernix K1  
C:Accession: F72471  
A:Reference: A72450; MUID:99310339  
A:Accession: F72471  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-133 <KAW>  
A:Cross-references: DDBJ:AP000064; NID:95105945; PIDN:BA81430.1; PID:dl045216; PID:95106119  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE2415  
F72471 Length: 133 February 11, 2000 15:52 Type: P Check: 430 ..

1 MLVKRDTSM LPTIATAPI ESRPMTTRAP VYALTALEM LLARFMPPT  
51 VAVAMIKPKV ADWTMCIMTS DILVATGA PSLATIPIMT RNPMLITISW  
101 RETGPITRSI CPISCRSGL GFLIKPGAPL RSR  
11AA\_SEQUENCE 1.0  
P1:C70405 - hypothetical protein aq\_1220 - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C>Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 08-May-1998  
C:Accession: C70405  
R:Accession: G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, R.; Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Feldman, D.E.; Overbeek, R.; Sneed, M.A.; Keller, M.; Aujay, M.; Huber, R.; Swanson, R.V. R.A.; Short, J.M.; Olson, G.J.; Swanson, R.V.  
Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666  
A:Accession: C70405  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-499 <AOE>  
A:Cross-references: GB:AE000729; NID:92983659; PID:92983669; GB:AE000657  
A:Experimental source: strain VFS  
C:Genetics:  
A:Gene: aq\_1220  
C70405 Length: 499 February 11, 2000 15:52 Type: P Check: 5299 ..

1 MMRKFLVLI PVILIFYNL GYNAVMKPN SEYDSAKNM LKSGFLIPV  
51 YNGEIRLNKP PMTYWISIG YKIGVNEIG LRFHALGV FTGVLTYLLA  
101 RRTGSKNTA LLSFLILIS FTIANARA SPEVPTFFI TLSLYMYEX  
151 FTRKRELFW LALIASLAV LKGPAGFVL PAGVVEFYLL LRAPEKLTKI  
201 KYAGILMWF LLSGMWELQ YIVAREELE VFKEKVKRI YALORDPEYF  
251 VALDIVSFL PYSLFFFLAL FWALKERRE LSPFLWFSF IFLIFSIVKM

301 KIPYINPAP PAMATITADF INSOLIKRKY NLSTIFIMY LVATIALSL  
351 YKFSATLFP LIPILLPFF LKRELLPAF GAFALFYLS SVILPYVOF  
401 RPYREVGKEI RLIDPKNELR TYLEGFEHNN LPEYADRVII RRTKEVKRPA  
451 IYLARKSFD CEPYRKWELY TSSESEFFKE MDIKRKRKF EEFILCVIK

11AA\_SEQUENCE 1.0  
P1:D70454 - hypothetical protein aq\_1793 - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C>Date: 08-May-1999 #sequence\_revision 08-May-1998 #text\_change 08-May-1998  
C:Accession: D70454  
R:Accession: G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, R.; Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Feldman, D.E.; Overbeek, R.; Sneed, M.A.; Keller, M.; Aujay, M.; Huber, R.; Swanson, R.V. R.A.; Short, J.M.; Olson, G.J.; Swanson, R.V.  
Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666  
A:Accession: D70454  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-269 <AOE>  
A:Cross-references: GB:AE000756; NID:92984076; PID:92984086; GB:AE000657  
A:Experimental source: strain VFS  
C:Genetics:  
A:Gene: aq\_1793  
D70454 Length: 269 February 11, 2000 15:52 Type: P Check: 3320 ..

1 MMSPMASIM YTHFRMKRHA LNWSTILFQ NFRLEAFELI LFTLLNELFL  
51 IPYNAVFAEF IHNVNFSLI IYFSKLYIKV KGNBEERYKRE IERTKLAQAL  
101 KYILPAVTL TEATYAMTA YILLFVSL ILGLINGITV FTFGIDPILIA  
151 YLIFIVILLI LYFWLITSYP VEPARTVIEG QTPNDEFFEL IMAFSPKLLW  
201 KLAFLSEVLF SSVYIGFESL FIFLFOFVMS HLFPPLFFEL YVAFSNTLL  
251 IYLFVGVISV YLWKRRECK  
11AA\_SEQUENCE 1.0  
P1:A75196 - hypothetical protein PAB0052 - Pyrococcus abyssi (strain Orsay)  
C:Species: Pyrococcus abyssi  
C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C:Accession: A75196  
R:Accession: Genoscope  
Submitted to the EMBL Data Library, July 1999  
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution.  
A:Reference number: A75001  
A:Accession: A75196  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-147 <KAW>  
A:Cross-references: GB:AJ248283; GB:AL096836; NID:95457433; PIDN:CA849016.1; PID:el514910; PID:95457525  
A:Experimental source: strain Orsay  
C:Genetics:  
A:Gene: PAB0052  
A75196 Length: 147 February 11, 2000 15:52 Type: P Check: 70 ..

1 MSSLSLAMP FLNSISGAT TLTSFIYAYF KAELITLINS LVLSGVSPS  
51 SLAKTINSF PPRASKREA TPISKPLAG IOTLIASLA FSKSSSSSLAV  
101 TSGTLPSTNL LPAIFNSCR KSITLIRKGT FMSLEKNSAT LFTYISS  
11AA\_SEQUENCE 1.0

```

P1:E75204 - hypothetical protein PAB0107 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: E75204
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: Insights into archaeal
chromosome structure and evolution.
A:Reference number: A75001
A:Accession: E75204
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-119 <KAM>
A:Cross-references: GB:A1248283; GB:AL096836; NID:g5457433; PIDN:CAB49084.1;
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB0107

E75204 Length: 179 February 11, 2000 15:52 Type: P Check: 8207 ..

1 MGYFKNKATN IKIRHIMKR DRGDFLSLS NPEISPIILL VIGSIFPDL
51 VTFYFSFESL ALHGEDIK EHRSHYHSL FLAPLVYSL AKSLFMFTI
101 GAASHLFLDF FSGVIFPFYP LRRKGYGVKI ILSIGTGFST KAKILRRPD
151 PRIERKIEIS KSIYLLTLT ILFLAKCK

!!NA_SEQUENCE 1.0
P1:H7185 - hypothetical protein PAB2346 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: H7185
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: Insights into archaeal
chromosome structure and evolution.
A:Reference number: A75001
A:Accession: H7185
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-266 <KAM>
A:Cross-references: GB:A1248283; GB:AL096836; NID:g5457433; PIDN:CAB48935.1;
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB2346

H7185 Length: 266 February 11, 2000 15:52 Type: P Check: 6767 ..

1 MIRVATFEDI DDMVSFIDA YNYTGRDSV AKSMELSLK OPNGCLIAFI
51 DCKPYGMGCI FLTKKAWMG LMGVKKVQR RGIETELFKR LKIKGKTI
101 RDDSAGQYG LYKKNFTDE YRTVRELRN RPLKKVSGV ELKEIPGWE
151 ELDKRAFQGD RTVRLIYLE SGARLIVEN EGSGMYRGR IGPLVAISRK
201 IAEKIMLKAF LLRGREIIVP DANEDAMLI KKSPEVLTS CYRMLGDKV
251 EENVHVFEGI LTVAKG

!!NA_SEQUENCE 1.0
P1:D71228 - hypothetical protein PH0086 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998
C:Accession: D71228
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto,
S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.;
Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohikku, Y.; Funahashi, T.;
Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushiida, N.; Oguchi, A.; Aoki, K.;
Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya,
H.; Kikuchi, H.
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a
hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.
A:Reference number: A71000; MUID:98344137
A:Accession: D71228
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-135 <KAM>
A:Cross-references: GB:AP000001; NID:g3236128; PID:d1030098; PID:g3256472
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by
GenBank
C:Genetics:
A:Gene: PH0086

D71228 Length: 135 February 11, 2000 15:52 Type: P Check: 1741 ..

1 MRGATTLTLE IVAYSKALLT LSTNSFVLSG KVSPPSLNRP TIISPPRAS
51 RREAILIKT KFLAGTHILL AVSLALSITS NSLAETLGTG PSTNALPESL
101 ENSLRKLLI IFKGLISFE KNSATLFTLL SLLP

!!NA_SEQUENCE 1.0
P1:D7110 - hypothetical protein PH0651 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998
C:Accession: D7110
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto,
S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.;
Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohikku, Y.; Funahashi, T.;
Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushiida, N.; Oguchi, A.; Aoki, K.;
Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya,
H.; Kikuchi, H.
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a
hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.
A:Reference number: A71000; MUID:98344137
A:Accession: D7110
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-102 <KAM>
A:Cross-references: GB:AP000003; NID:g3236130; PID:d1030685; PID:g3257059
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by
GenBank
C:Genetics:
A:Gene: PH0651

D7110 Length: 102 February 11, 2000 15:52 Type: P Check: 6672 ..

1 MLTISLGM NIAKEFNVVY LPEVPATR MFAGLTPRS TRSQNATS
51 EFKVNFIS IIVSGSLNF LIVVGPRL IGSNVALTLL PSQLASMSG
101 FC

!!NA_SEQUENCE 1.0
P1:E71120 - hypothetical protein PH0732 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998
C:Accession: E71120
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto,
S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.;
Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohikku, Y.; Funahashi, T.;
Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushiida, N.; Oguchi, A.; Aoki, K.;
Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya,
H.; Kikuchi, H.
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a
hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.
A:Reference number: A71000; MUID:98344137

```



Mon Feb 14 08:07:32 2000

pir.cat

Page 189

A:Accession: E71120  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-104 <KAM>  
A:Cross-references: GB:AP000003; NID:g3236130; PID:d1030766; PID:g3257140  
A:Experimental source: strain OT3  
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C:Genetics:  
A:Gene: PH0732

E71120 Length: 104 February 11, 2000 15:52 Type: P Check: 8797 ..

1 MAHFNNHYF PRLHCHYAI PINEGMHYI VRSSPRDLSC YQFISIRYVS  
51 KAKISYLLNS CLLKSTIYI HSTSKNEMV NLYHRNVLN TFSYNGSFS  
101 FINI

!!AA\_SEQUENCE 1.0  
P1:H71124 - hypothetical protein PH0767 - Pyrococcus horikoshii  
C:Species: Pyrococcus horikoshii  
C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 14-Aug-1998  
C:Accession: H71124  
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.  
DNA Res. 5, 55-76, 1998  
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.  
A:Reference number: A71000; MUID:98344137  
A:Accession: H71124  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-272 <KAM>  
A:Cross-references: GB:AP000003; NID:g3236130; PID:d1030801; PID:g3257175  
A:Experimental source: strain OT3  
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C:Genetics:  
A:Gene: PH0767

H71124 Length: 272 February 11, 2000 15:52 Type: P Check: 5357 ..

1 MANFNRLSS SSAVTRPPE ALVIMATLP LGKGCENALA MTSISIVFA  
51 TKTPPLCINAA SYIPOLPAKE PVMLEALAP LSLPFAEKAM IGFEVTLFA  
101 TSMNFFPFLT PSNRYAIALS SSVSAKYSIS SAKSTSAMLP TLNFPENPMS  
151 SEIVLYITPA PNAPLMINNA TPEPLATVIK VAFKSTWVFI IPTQFPMIL  
201 IPIYMAISTS SFSLEPSSP VSPNPVITIT TPLTPSPHW ISVSFTNLFG  
251 TTTIAKSSS GISLTLLVAL TP

!!AA\_SEQUENCE 1.0  
P1:F71127 - hypothetical protein PH0787 - Pyrococcus horikoshii  
C:Species: Pyrococcus horikoshii  
C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 14-Aug-1998  
C:Accession: F71127  
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.  
DNA Res. 5, 55-76, 1998  
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.

A:Reference number: A71000; MUID:98344137  
A:Accession: F71127  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-183 <KAM>  
A:Cross-references: GB:AP000003; NID:g3236130; PID:d1030823; PID:g3257197  
A:Experimental source: strain OT3  
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C:Genetics:  
A:Gene: PH0787

F71127 Length: 183 February 11, 2000 15:52 Type: P Check: 8012 ..

1 MLILHSEGA SVEDCIRTLI RMSKNRLPQ GKITSKIKI STGAFSLVNL  
51 ALIVDLAPY KPGIAVEYSV GSKRKALED LQELNSVYT PEIEVFDFOI  
101 EYTPVTRR TYAIGLVYN KPRANKDF MLONRKILA KYLELLVNI  
151 KALNISELAR MEGVSRDITV NDIOQIIKNV DKV

!!AA\_SEQUENCE 1.0  
P1:F71039 - hypothetical protein PH1606 - Pyrococcus horikoshii  
C:Species: Pyrococcus horikoshii  
C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 14-Aug-1998  
C:Accession: F71039  
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.  
DNA Res. 5, 55-76, 1998  
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.  
A:Reference number: A71000; MUID:98344137  
A:Accession: F71039  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-447 <KAM>  
A:Cross-references: GB:AP000006; NID:g3236133; PID:d1031661; PID:g3258035  
A:Experimental source: strain OT3  
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C:Genetics:  
A:Gene: PH1606

F71039 Length: 447 February 11, 2000 15:52 Type: P Check: 7274 ..

1 MKKIKSGII ISALYTICML YLTQYINGL RYRSDLGIF AQSLSLSGG  
51 KLFYNTVERO LYNVOHFEV HPQPLELLE PLFLFKKT YALVVSOTIA  
101 LGTSVYLAVR LATELGEKK GVALTILYAC NSSLIGINIF EHPVSLAVP  
151 LFLAVKFK RSPLEFIIS TLILTKEDA FLVVISIIV KILKEGFEVE  
201 ILKRNKRIIF FAPLALVGI LITKYIPEK GKRIYISIV TRFSIDRRKL  
251 LYELLENTLF ALPEMLDFG MSLIPWLE CLASRETOI MGFHYPMIL  
301 VPLSEVASLE VAROLKRIY KRVITIGILA SLATLPIAKL PEKPNPLIY  
351 LVTITPVGK EASWEALIV KGLKPIYTO PEYFPLAIR SDVYIYPKNV  
401 NRVILVNLN TYGRRALKR LRDFKVELR YKKIEKDV IIMLRD

!!AA\_SEQUENCE 1.0  
P1:F71201 - hypothetical protein PH1883 - Pyrococcus horikoshii  
C:Species: Pyrococcus horikoshii  
C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 14-Aug-1998  
C:Accession: F71201

R.Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, O.; Sekine, M.; Babu, S.; Kosugi, H.; Hosoyama, A.; Nagai, T.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohfuku, T.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushiida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.

A:Title: Complete sequence and gene organization of the genome of a hyperthermophilic archaeobacterium, *Pyrococcus horikoshii* OT3.  
A:Reference number: A71000; M01D:98344137  
A:Accession: F71201  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-137 <KAM>  
A:Cross-references: GB:AP000007; NID:93236134; PID:d1031948; PID:93258322  
A:Experimental source: strain OT3  
A:Note: this accession replaces an interim accession for a sequence replaced by C:Genetics:  
A:Gene: PH1883

F71201 Length: 137 February 11, 2000 15:52 Type: P Check: 6730 ..

1 MSAYITGSL SVSONSESIP CGAVSNMPP PLFLMSKQL VSIPSKNIGS  
51 FTYSEIYMY SSPISPLSI ISLAFCSGW NLGKGTISL TPAFLAFTI  
101 SLPSSVSAM GFSRSISFPS SMAIATSLC LYVSIM

!!AA\_SEQUENCE 1.0  
P1:D71216 - hypothetical protein PH196 - *Pyrococcus horikoshii*  
C:Species: *Pyrococcus horikoshii*  
C:Date: 14-Aug-1998 #sequence-revision 14-Aug-1998 #text-change 14-Aug-1998  
A:Accession: D71216  
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, O.; Sekine, M.; Babu, S.; Kosugi, H.; Hosoyama, A.; Nagai, T.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushiida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.

DNA Res. 5, 55-76, 1998  
A:Title: Complete sequence and gene organization of the genome of a hyperthermophilic archaeobacterium, *Pyrococcus horikoshii* OT3.  
A:Reference number: A71000; M01D:98344137  
A:Accession: D71216  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-225 <KAM>  
A:Cross-references: GB:AP000007; NID:93236134; PID:d1032066; PID:93258440  
A:Experimental source: strain OT3  
A:Note: this accession replaces an interim accession for a sequence replaced by C:Genetics:  
A:Gene: PH196

D71216 Length: 225 February 11, 2000 15:52 Type: P Check: 2646 ..

1 MCSFLMIAF DPGTLVDSY SCIEAFYFA LRRTYWPFG KRYMKLLTK  
51 IELQFERPRF GRHSGRIKAP MKIFQGFAR IWEERAKLT KELEGAREVL  
101 KELKREGHIV ISFSADFTF GYKEKRLAS GLYDIEDVLI IGHGVSIKE  
151 AFRIVREKYG NETFWVWDK PMRFIGNDE NTEYVMWTF YTAPEVTREI  
201 LDKIPHLHVI YDLMSLDVL RLRSY

!!AA\_SEQUENCE 1.0  
P1:G75035 - Iyon (iii) abc transporter, permease protein (hemu-2) PAB1535 - *Pyrococcus abyssi* (strain Orsay)  
C:Species: *Pyrococcus abyssi*  
C:Date: 20-Aug-1999 #sequence-revision 20-Aug-1999 #text-change 20-Aug-1999  
A:Accession: G75035

R:anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999  
A:Description: *Pyrococcus abyssi* genome sequence: insights into archaeal chromosome structure and evolution.  
A:Reference number: A75001  
A:Accession: G75035  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-331 <KAM>  
A:Cross-references: GB:AJ248287; GB:AL096836; NID:95458657; PIDN:CAB50180.1;  
A:Experimental source: strain Orsay  
C:Genetics:  
A:Gene: PAB1535

G75035 Length: 331 February 11, 2000 15:52 Type: P Check: 3225 ..

1 MKKVTALIL ASITSLYLAL SLGSVRIPLS TLINSLSH ISLYRGOLS  
51 GSPYIIIGIR LPRVMAIYV GFSIALAGTA SOALEFNPPLA DPYIIIGISG  
101 ASIGAAIATA YSPRYVEYFA FVGAIMAVYL VYRISKVNH IPVVLIIAG  
151 IAVGFESHAV TSYILXNRD KVHGLSWLF GTLALATMSK VFIMAVAVGI  
201 GGGMLFMSR ELNLLLGE E SIALGDLVNL YRLLIIFALA ILTGVAVSES  
251 GIIGFIIGLVS PHIMRMFVG P NHRLPVS A MUGILVLVIS DLISRTIVSP  
301 VEIPGVIVTA LFGAFPFAYL LMRKRGELY A

!!AA\_SEQUENCE 1.0  
P1:B75110 - lipote-protein ligase A related PAB1916 - *Pyrococcus abyssi* (strain Orsay)  
C:Species: *Pyrococcus abyssi*  
C:Date: 20-Aug-1999 #sequence-revision 20-Aug-1999 #text-change 20-Aug-1999  
A:Accession: B75110  
R:anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999  
A:Description: *Pyrococcus abyssi* genome sequence: insights into archaeal chromosome structure and evolution.  
A:Reference number: A75001  
A:Accession: B75110  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-249 <KAM>  
A:Cross-references: GB:AJ248285; GB:AL096836; NID:95458067; PIDN:CAB49595.1;  
A:Experimental source: strain Orsay  
C:Genetics:  
A:Gene: PAB1916

B75110 Length: 249 February 11, 2000 15:52 Type: P Check: 5732 ..

1 MRFPIIVAR PELQMAIDEA ILIARSGKV PPTVALYFK PSSVTIGRFQ  
51 SVRHVDIER ARELNIPYVR RITGGGSYFH DEFGITTSV VISDDYHPST  
101 KNIOESYRFL AGPLVDALKD LGINAFPSGI NDIYVNGKI SGAQOTFRKG  
151 IILQGTIFY STRLEVLAIV LKYSKRLKD KGISIYERV TTIEREGIKL  
201 SMOETIELR NSFRRAPLE EGELTYELE LAQELVEERY GKDEWFMK

!!AA\_SEQUENCE 1.0  
P1:E49094 - methylmalonyl-CoA decarboxylase (EC 4.1.1.41) beta chain - *Veillonella parvula*  
C:Species: *Veillonella parvula*  
C:Date: 19-Mar-1997 #sequence-revision 19-Mar-1997 #text-change 26-Feb-1998  
A:Accession: E49094  
R:Huder, J.B.; Dimroth, P.  
J. Biol. Chem. 268, 24564-24571, 1993  
A:Title: Sequence of the sodium ion pump methylmalonyl-CoA decarboxylase from

Mon Feb 14 08:07:32 2000

pir.cat

Page 191

Veillonella parvula  
A:Reference number: A49094; MUID:94043308  
A:Accession: F49094  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1373 <HND>  
A:Cross-references: GB:L22208; NID:9415592; PID:9415597  
C:Keywords: carbon-carbon lyase; carboxy-lyase  
E49094 Length: 373 February 11, 2000 15:52 Type: P Check: 3704 ..  
1 MEAFVAIAOS VINDSGFLAF TTGNAMILV GILLYLAFA REEPPLIGP  
51 IAFGLLANI PRNGFEEGYV ALISGISOE IEPPLIFIGV GAMTDFGPI  
101 ANPKTLILGA AAOIGVPAAL GGAMMLGFTA QEAAAGIIG GADGPTIYL  
151 ATRAPHLIG AIAVAASYM SLVPLIQPV MKLFTQKER EIVMEQLREV  
201 TREKIVEPI VAIIFISLL PSITSLGLM MGLNFEREG VTRLDSTSO  
251 NALINTVITF LATGTGLMS AEHFLSLEI KIIILGLFAP IGTAGVLF  
301 GKILSLVDGG KTNPLIGSAG VSAVPMARV SQVGAKANP ANFLMHANG  
351 PNVAAGVIGA VAAGTMLAML SNR  
!!AA\_SEQUENCE 1.0  
P1:14222 - NADH dehydrogenase subunit II - Xantusia vigillis mitochondrion  
(Scc1)  
C:Species: mitochondrion Xantusia vigillis  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T14222  
R:Macvey, J.R.; Larson, A.; Anandjaya, N.B.; Fang, Z.; Papenfuss, T.J.  
Mol. Biol. Evol. 14, 91-104, 1997  
A:Title: Two novel gene orders and the role of light-strand replication in  
rearrangement of the vertebrate mitochondrial genome.  
A:Reference number: 217789  
A:Accession: T14222  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1344 <MAC>  
A:Cross-references: EMBL:U71328; NID:91753276; PID:91753278; PIDN:ABA8290.1  
C:Genetics:  
A:genome: mitochondrion  
A:genetic code: Scc1  
C:Note: ND2  
C:Keywords: mitochondrion  
T14222 Length: 344 February 11, 2000 15:52 Type: P Check: 7879 ..  
1 MSPTIOSLIL SSIATGIIIT MTSQSWILAM VOLEINTLAI LPIISPHHP  
51 RATEAMTKYF LQOTAASTMI LPASTLNAMH SGOMINMOLT TEPASTVLT  
101 ALTKKLGLAP VHFVLPYVQO GSTINTAMIL TWOKLADMT ILIMTHQNP  
151 TTIILITGII SSIMSGMAGL NQOLRKILA YSSIAHLGMI YVAISDETI  
201 TITLLIYIL LSLIFLSMN TSHKTKMDM ATTWISLPTL TTEMILTLIS  
251 LGGLPPLNG LPRWILQEL TTHHTPIAT ILAFSLISL FFYRLSYIT  
301 TITLHPNPT MSNKWRCKY KSKSLISTMM FVTTMIPMS PIML  
!!AA\_SEQUENCE 1.0  
P1:S55472 - pectin lyase (EC 4.2.2.10) - Mycosphaerella pinodes  
C:Species: Mycosphaerella pinodes  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 10-Sep-1997  
C:Accession: S55472  
R:Heim, P.  
Submitted to the EMBL Data Library, May 1995  
A:Reference number: S55472

A:Accession: S55472  
A:Molecule type: DNA  
A:Residues: 1299 <HEI>  
A:Cross-references: EMBL:X87580; NID:9854337; PID:9854338  
A:Experimental source: strain Berkeley and Bloxom  
C:Genetics:  
A:gene: pella  
A:introns: 43/3  
C:Keywords: carbon-oxygen lyase  
S55472 Length: 299 February 11, 2000 15:52 Type: P Check: 8743 ..  
1 MKEFVSIG LTATLALAP TSMIDVARD VKRASLED VATAERPLRC  
51 RPSASLPSPS RYTRLPPLSS LVLSRALATT SRSAATSPSS DTRASCLRT  
101 SPSPSRVSRM LSFVTLPSR SSVAMLSPSR RYELVTPKEI TYLHRLQSE  
151 NWVDHILDS SPRDNDKRY DGLDITHAH DFVTVITFL HDHMKASLIG  
201 HSDNGAEDK GHLYTYANN YLKNUNSRG PSASQATST TTTTTCRMV  
251 STPARVSCIL FRTTLASPR RLTRRLMDV LLSLATSVM ARTLPRVL  
!!AA\_SEQUENCE 1.0  
P1:D72485 - probable high-affinity branched-chain amino acid transport  
ATP-binding protein APE2524 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C:Accession: D72485  
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,  
R.; Kawarabayashi, M.; Sekine, M.; Baba, S.; Ankel, A.; Hosoyama, A.;  
Fuku, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamida, M.; Masuda, S.;  
Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kishida, N.; Oguchi, A.;  
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.  
DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic  
Crenarchaeon, Aeropyrum pernix K1.  
A:Reference number: A72450; MUID:99310339  
A:Accession: D72485  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1242 <RAW>  
A:Cross-references: DDBJ:AF000064; NID:95105945; PIDN:BAAB1540.1; PID:dl045326;  
PID:95106229  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE2524  
D72485 Length: 242 February 11, 2000 15:52 Type: P Check: 7054 ..  
1 MEVITILDV RDYKRRFGI EALKVSFSV RRGERRGLIG PNGAGKTLTF  
51 NIISGIYMPD RGRVYIKGYD ITGMPAYRRS RAGIARTFOI VRPLANTVL  
101 NNVIYGALIR TNDIRERAR AMEADIMVGL AGKEDILAND LNLIEKRLE  
151 LARALATQPE LLLIDELIAG LRPREVDDLV YTLLEISKRG ITIIMEVHM  
201 RAVMFAERV IVLHFGKIA EGTPREVASN KLVIEAYNGI GG  
!!AA\_SEQUENCE 1.0  
P1:B71024 - probable lipate protein ligase - Pyrococcus horikoshii  
C:Species: Pyrococcus horikoshii  
C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 14-Aug-1998  
C:Accession: B71024  
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto,  
S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.;  
Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamida, M.; Ohfuku, Y.; Funahashi, T.;  
Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kishida, N.; Oguchi, A.; Aoki, K.;  
Yoshizawa, T.; Nakamura, Y.; Kobb, F.T.; Horikoshi, K.; Maechi, Y.; Shizuya,  
H.; Kikuchi, H.  
DNA Res. 5, 55-76, 1998

A>Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, Pyrococcus horikoshii OT3.  
A:Reference number: A71000; MUID:98344137  
A:Accession: B71024  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-249 <KAW>  
A:Cross-references: GB:AP000006; NID:g3236133; PID:d1031537; PID:g3257911  
A:Experimental source: strain OT3  
A>Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C:Genetics:  
A:Gene: PH1487  
B71024 Length: 249 February 11, 2000 15:52 Type: P Check: 6905 ..  
1 MRFPILVAR PEVQMAIDEA ILIARSEGRV PDTVRLYFK PSSVTIGRFQ  
51 SVRYDNLEK AKELGIPYVR RITGGGSVFH DEYGEITYSV VISEDYHPML  
101 KNQDSYRFL AGPLVDALKD LGIEAEFSLG NDIYVRGKTI SGAQOTRRRG  
151 ITHQHTLMY ATRLEILGSV LKYSKEKLD KGVKSIPEPV TILEREGIKY  
201 GRNEMTEVLN NSFRAFLPD EGELETELE LVEKLIBERY GNDKMFQK  
11AA\_SEQUENCE 1.0  
P1:572288 - ribosomal protein l23 - Plasmodium falciparum plastid  
C:Species: Plasmid Plasmodium falciparum  
C>Date: 14-Apr-1998 #sequence\_revision 24-Apr-1998 #text\_change 07-May-1999  
C:Accession: S72288  
R:Wilson, R.J.M.; Denny, P.W.; Preiser, P.R.; Rengschart, K.; Roberts, K.; Roy, J.; Whyte, A.; Strach, M.; Moore, D.D.; Moore, P.W.; Williamson, D.H.  
A>Title: Complete gene map of the plastid-like DNA of the malaria parasite Plasmodium falciparum.  
A:Reference number: S72277; MUID:96346169  
A:Accession: S72288  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-75 <WTL>  
A:Cross-references: EMBL:X55276; NID:g1171591; PID:e220177; PID:g1171594  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1998  
C:Genetics:  
A:Gene: rpl23  
A:Genome: plastid  
A>Note: this apparently degenerate plastid is referred to as the apicoplast  
C:Keywords: plastid; protein biosynthesis; ribosome  
S72288 Length: 75 February 11, 2000 15:52 Type: P Check: 9584 ..  
1 MKEVILNPLI YNLFYKINV LNKFCITYSI KFTFLDIKY IINKIRIKI  
51 INTNIRIKIN INKNCKLAKY YITFK  
11AA\_SEQUENCE 1.0  
P1:110349 - Very-late factor 1 protein - Oryza pseudotsugata nuclear polyhedrosis virus  
C:Species: Oryza pseudotsugata nuclear polyhedrosis virus, OPMNPV  
C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 16-Jul-1999  
C:Accession: T10349  
R:Ahrens, C.A.; Russell, R.R.; Funk, C.J.; Evans, J.; Harwood, S.; Rohrmann, G.F.  
A>Title: The sequence of the Oryza pseudotsugata multinucleocapsid nuclear polyhedrosis virus genome.  
A:Reference number: Z17011  
A:Accession: T10349  
A>Status: preliminary; translated from GB/EMBL/DBD  
A:Molecule type: DNA  
A:Residues: 1-374 <AHR>  
A:Cross-references: EMBL:U75930; NID:g2934903; PID:g1911326  
T10349 Length: 374 February 11, 2000 15:52 Type: P Check: 889 ..  
1 MDGLGVANET TENDWKARIQ SASRFEHVD IATDRQRTCP DEVKNDLSWS  
51 KYEPKPEFAP TLKSYKSRLL IKIIFSLVEE ADLQNPAYDL DREFDSEYEQ  
101 HLAVSPRELQ KMLRELRSVT KETLQTLIN YTNANLPEEF KIPRWMLPR  
151 DKELKIREK EKNFKLNNAI DTILNFDISK ILMNSDGVH DRGLINGAIV  
201 FCIMLGTGMR IENARQSYD DLNVLKKGK LRSPTIGLKR KRSRKNLTNN  
251 IKTRPLELAR EYARNPVL QISKNTSTPF KDFRLLDEA GVEMERPSN  
301 MIRHYLSSNL YNSGVPLQKV ARLNHESPA STKPYLNRYN FDESSDEES  
351 GGNNRDSSTG SSANSSSLY QTCG  
11AA\_SEQUENCE 1.0  
P1:578187 - ymf39 protein - Reclinomonas americana (ATCC 50394) mitochondrion  
C:Species: mitochondrion Reclinomonas americana  
A:Variety: ATCC 50394  
C>Date: 29-Jan-1998 #sequence\_revision 06-Feb-1998 #text\_change 17-Mar-1999  
C:Accession: S78187  
R:Lang, B.F.; Burger, G.; O'Kelly, C.J.; Cedergren, R.; Golding, G.B.; Lemieux, C.; Sankoff, D.; Turmel, M.; Gray, M.W.  
A>Title: An ancestral mitochondrial DNA resembling a eubacterial genome in mitature.  
A:Reference number: S78127; MUID:97311393  
A:Accession: S78187  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-197 <LAN>  
A:Cross-references: EMBL:AF007261; NID:g2258325; PID:g2258386  
A:Experimental source: ATCC 50394  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1997  
C:Genetics:  
A:Gene: ymf39  
A:Genome: mitochondrion  
C:Keywords: mitochondrion  
S78187 Length: 197 February 11, 2000 15:52 Type: P Check: 3067 ..  
1 MLSEIQKHV TLTLLVPLVA ILARDEIV HDETLVLCF ILFLTFLYVA  
51 LKDWYASFN DRALQIERF NDSYSLEQOT LQLLANHEK QVSLNLEIDS  
101 LMEFTKSEVN NIQTRQAL RARLISEFT KNTAVKKED AFOFIQOTT  
151 NTIYANISLE NISGPTGEL KNTSFEGR ILDEBAISLQ DKREANS  
11AA\_SEQUENCE 1.0  
P1:111851 - ruvB Protein - Thermus aquaticus thermophilus  
C:Species: Thermus aquaticus thermophilus  
C>Date: 14-Aug-1999 #sequence\_revision 14-Aug-1999 #text\_change 14-Aug-1999  
C:Accession: T11851  
R:Tong, J.; Wetmur, J.G.  
A>Title: Cloning, sequencing, and expression of ruvB and characterization of RuvB proteins from two distantly related thermophilic eubacteria.  
A:Reference number: Z17360  
A:Accession: T11851  
A>Status: preliminary  
A:Residues: 1-324 <TON>  
A:Cross-references: EMBL:U22817; NID:g727372; PID:g727373  
T11851 Length: 324 February 11, 2000 15:52 Type: P Check: 3584 ..  
1 VEDALRPKT LDEYGOERL KOKLRVYLA AKARREPLEH LLFGPGLG

Mon Feb 14 08:07:32 2000

pir.cat

Page 193

51 KTLAHVIAH ELGVNLRVTS GPALEKPGDL AAILANSLEE GDILFIDEIH  
101 RLSROAEHL YPAMEDFVMD IVIGGPAAR TIRLEPRFA LIGATRPGL  
151 ITADILSRFG IVEHLEYTP BELAQVARD ARLGVRITE EAALEIGRS  
201 RGTMRVAKRL FRRVRDAQV EGEEVITRR ALEALAAGL DELGLEKDR  
251 EILEVILIRE GAGPVGLATL ATALSEDPT LEEVHEPYLI RQGLKRTPR  
301 GRATELAVR HLGYPVPV LLEP

!!AA-SEQUENCE 1.0  
P1:JC5705 - alpha-amylase inhibitor 0.19 - synthetic  
C:Species: synthetic  
A:Note: Triticum aestivum (common wheat) gene engineered and expressed in  
Escherichia coli  
C>Date: 03-Sep-1998 #sequence\_revision 03-Sep-1998 #text\_change 03-Sep-1998  
C:Accession: JC5705  
R:Okuda, M.; Satoh, T.; Sakurai, N.; Shibuya, K.; Kajii, H.; Samejima, T.  
J. Biochem. 122, 918-926, 1997  
A:Title: Overexpression in Escherichia coli of chemically synthesized gene for  
active 0.19 alpha-amylase inhibitor from wheat kernel.  
A:Reference number: JC5705; MUID:98104043  
A:Accession: JC5705  
A:Molecule type: DNA  
A:Residues: 1-125 <OK>  
A:Cross-references: DBJ:AB003682; NID:92575815; PID:41020955; PID:92116581  
C:Comment: The sequence from wheat (see PIR:A01324) has signal and propeptide  
sequences not included in this synthetic sequence.

JC5705 Length: 125 February 11, 2000 15:52 Type: P Check: 3534 ..

1 MSGPMCTPG QAFQVPALPA CRPLRLQCN GSGVPEAVLR DCCQOLAHIS  
51 EMCRCGALIS MLDMSYKEHG AOGGAGTGA FPRCRREYVK LTAASITAVC  
101 RLPIVDASG DGAIVCKDVA AYPDA

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100

1

2

3

4

5

6

7

8

9

10

11

12

13